



## King's Research Portal

DOI:

[10.1112/plms.12226](https://doi.org/10.1112/plms.12226)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Armstrong, J., Brigo, D., & Rossi Ferrucci, E. (2019). Optimal approximation of SDEs on submanifolds: the Ito-vector and Ito-jet projections. *PROCEEDINGS OF THE LONDON MATHEMATICAL SOCIETY*, 119(1), 176-213. <https://doi.org/10.1112/plms.12226>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Optimal approximation of SDEs on submanifolds: the Itô-vector and Itô-jet projections

John Armstrong, Damiano Brigo and Emilio Rossi Ferrucci

## ABSTRACT

We define two new notions of projection of a stochastic differential equation (SDE) onto a submanifold: the Itô-vector and Itô-jet projections. This allows one to systematically develop low dimensional approximations to high dimensional SDEs using differential geometric techniques. The approach generalizes the notion of projecting a vector field onto a submanifold in order to derive approximations to ordinary differential equations, and improves the previous Stratonovich projection method by adding optimality analysis and results. Indeed, just as in the case of ordinary projection, our definitions of projection are based on optimality arguments and give in a well-defined sense “optimal” approximations to the original SDE in the mean-square sense over small times. We also explain how the Stratonovich projection satisfies an optimality criterion that is more ad hoc and less appealing than the criteria satisfied by the Itô projections we introduce.

As an application, we consider approximating the solution of the non-linear filtering problem with a Gaussian distribution. We show how the newly introduced Itô projections lead to optimal approximations in the Gaussian family and briefly discuss the optimal approximation for more general families of distributions. We perform a numerical comparison of our optimally approximated filter with the classical Extended Kalman Filter to demonstrate the efficacy of the approach.

## 1. Introduction

In this paper we define three notions of projecting a stochastic differential equation (SDE) onto a (sub)manifold  $M$ . Our aim is to derive practical numerical methods for solving SDEs and we will illustrate our theory with an example drawn from signal processing.

To explain the general idea, let us first consider projecting an ordinary differential equation (ODE) from the Euclidean space  $\mathbb{R}^r$  onto an  $n$ -dimensional manifold  $M \subseteq \mathbb{R}^r$ ,  $n < r$ . An ODE in  $\mathbb{R}^r$  can be thought of as defining a vector field in  $\mathbb{R}^r$ . At every point  $x \in M$  we can use the Euclidean metric to project the vector at  $x$  onto the tangent space  $T_x M$ . In this way one obtains a vector field on  $M$  which can be thought of as a new ODE on  $M$  that approximates the full ODE in  $\mathbb{R}^r$ . This is illustrated in Figure 1. It is easy to prove that this will be the best way of approximating the ODE in  $\mathbb{R}^r$  with an ODE on  $M$ . To be precise, if the initial condition for an ODE is a point  $x$  on the manifold, then any curve on  $M$  with tangent not equal to the projected vector field will diverge from the solution to the ODE faster than a curve which is tangent to the projected vector field. In this sense, the projected ODE is the only ODE which is asymptotically “optimal” at each point for small times. This paper addresses the question of how projection can be generalized from ODEs to SDEs. After some brief preliminaries on Itô–Taylor series in Section 2, and on the jet formulation of SDEs on manifolds in Section 3 we answer this question by describing three possible generalizations to SDEs in Section 4.

The first generalization of projection to SDEs has been proposed previously: what we shall call the Stratonovich projection. The Stratonovich projection is obtained by simply applying the

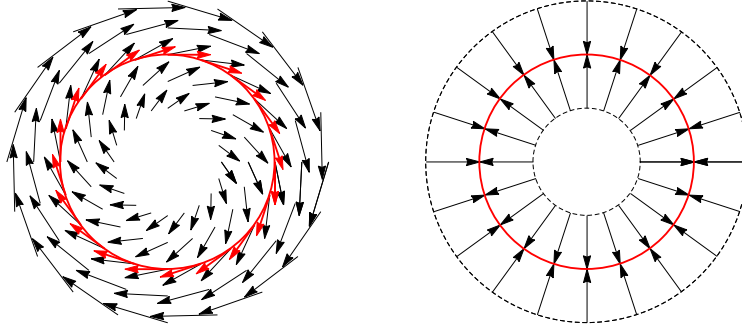


FIGURE 1. *Left: A pictorial representation of the projection of an ODE defined on  $\mathbb{R}^2$  to an ODE defined on the circle. Right: Metric projection from  $\mathbb{R}^2$  to the circle.*

projection operator to the coefficients of the SDE written in Stratonovich form. This projection has simply been derived heuristically from the deterministic case. Nevertheless, it appears to be a good approximation in practice and it has been used to find good quality numerical solutions to the non-linear filtering problem (See [10], [11], [5]). The Stratonovich projection is a natural first choice from the following point of view. As is obvious to anyone with experience of stochastic differential equations on manifolds, and we refer to the monographs and articles [17], [15], [19], [14], [21], [3], simply applying the projection operator to the coefficients of the SDE written in Itô form will not work. This is because solutions to the projected equation don't stay on the manifold, contrary to the Stratonovich case. Nevertheless, we will be able to obtain two modifications of this idea, which we will call the *Itô-vector projection* and the *Itô-jet projection*. These both give well-defined SDEs on the manifold.

We derive the Itô-vector projection by seeking an SDE on the manifold which optimally approximates the original SDE on the manifold over small times when the size of the errors are measured in the mean square ambient metric of  $\mathbb{R}^r$ . The mean squared error between a trajectory following the original SDE and following an SDE on the manifold will typically grow at a rate  $O(t^{\frac{1}{2}})$ . The diffusion term of the projected SDE is determined by minimizing the coefficient of  $t^{\frac{1}{2}}$  in this growth estimate. Choosing the drift term is more delicate, but we give two minimization arguments that indicate that the optimal choice of drift term is given by what we call the Itô-vector projection. The first argument identifies the drift by minimizing the coefficient of the  $O(t)$  term in the estimate of the error, notwithstanding the fact that there is also an  $O(t^{\frac{1}{2}})$  term. The second argument is to find an SDE on the manifold such that the difference between the means of the solutions to the original SDE and to the SDE on the manifold are minimized.

Both of these arguments are somewhat unsatisfying. As an alternative approach we consider finding the SDE on the manifold that most closely tracks the *metric projection* of the solution to the original SDE. The metric projection is the map that sends a point in the ambient space to the closest point of an embedded manifold  $M$ . It is well known to be well-defined and smooth on a tubular neighbourhood. The metric projection is illustrated in Figure 1. It is possible to find an SDE on the manifold such that the mean squared distance between the solutions on the manifold and the metric projection of the solution to the original SDE grows at a rate  $O(t)$ . This requirement determines the diffusion term of the SDE on the manifold and makes the  $O(t^{\frac{1}{2}})$  term coefficient vanish, rather than merely minimize it. Minimizing the coefficient of the order  $t$  term in this estimate determines the drift. We call the SDE determined in this way the *Itô-jet projection*.

It is natural to ask if the Stratonovich projection can also be derived from an optimality argument. We will explain that the Stratonovich projection is optimal when using a time-reflection-symmetric optimality criterion anchored to the deterministic initial condition of the process as a special state. However, as we will see, for our applications to filtering, the form of optimality achieved by the Stratonovich projection is not particularly useful. This is because the filtering problem is inherently asymmetric in time, as indeed are most applications of SDEs. Nevertheless, it is conceivable that in some applications of SDEs to physics, time reversal symmetry may be a paramount concern. In this case the Stratonovich projection may be preferred.

Surprisingly the Itô-vector projection, the Itô-jet projection and the Stratonovich projection are all distinct. All of them reduce to classical projection in the case of ODEs. Thus, while optimality arguments lead to a single best method for projecting ODEs, the situation is more complex for SDEs. Since both Itô projections are derived from optimality arguments that are much less ad hoc than the argument for optimality of the Stratonovich projection, there is a clear sense in which they are an improvement upon the Stratonovich projection—both theoretically and in practice.

However, it is not immediately clear whether one should prefer the Itô-vector or the Itô-jet projection. We investigate this question in Section 5, in which we consider a simple toy example which we believe strongly suggests that the Itô-jet projection is the better approximation. We also prove a simple theorem that shows how this example can be generalized.

We use this same toy example to illustrate another (entirely non-rigorous) reason for preferring the Itô-jet projection: mathematical aesthetics. As we shall see, each of the different notions of projection is best understood using different formulations of SDEs on manifolds. As its name suggests, the Stratonovich projection is most readily understood using Stratonovich calculus. The Itô-vector projection is most readily understood using the formulation of SDEs on manifolds in terms of Itô calculus first introduced by Itô in [22]. Finally, the Itô-jet projection is most readily understood using the 2-jet formulation of [3]. As we will see, the Itô-jet projection has a very elegant formulation in the language of 2-jets. It is even possible to draw a diagram that allows one to interpret the Itô-jet projection visually. We will present a diagram that visually represents the Itô-jet projection of our toy example. In fact, the development of the 2-jet formulation of SDEs in [3] was originally motivated by the development of these projection methods. It is for this reason that we have called the projections the Itô-vector and Itô-jet projections respectively.

Section 6 is devoted to a detailed calculation of the Itô-jet projection in local coordinates. This calculation amounts to computing the Taylor series for the metric projection map up to second order, and is essential to using the projection for applications.

Section 7 demonstrates how the notion of projection can be applied in practice. In particular, we will apply it to the non-linear filtering problem. We will derive general projection formulae for the non-linear filtering problem. We will then apply this to the problem of approximating a non-linear filter using a Gaussian distribution. A reader who is unfamiliar with non-linear filtering will want to consult Section 7.1 for a brief review.

Gaussian approximations to non-linear filters are widely used in practice (see for example [23, 8]). In particular, the Extended Kalman Filter (EKF) is a popular approximation technique. Other Gaussian approximations exist such as Assumed Density Filters (ADF) and filters derived from the Stratonovich projection. Our theory indicates that all these classical techniques can be improved upon by using the Itô projections (at least over small time intervals). We confirm this with a numerical example.

The utility of the projection method is by no means restricted to the filtering problem nor to such simple approximations as Gaussian filters. Our previous work [7] shows how the Stratonovich projection can be used to generate far more sophisticated filters and it is clear that the idea of projection should be widely applicable in the study of ODEs, SDEs, PDEs and

SPDEs. Nevertheless by focussing on Gaussian filters we can examine in detail the idea that there may be many useful ways of approximating an SDE on a submanifold, but that the Itô projections are in some sense optimal amongst these approximation methods. The point we wish to emphasize is that the Itô projections are able to tell us something new even about the well-worn topic of approximating the non-linear filtering problem using Gaussian distributions.

Finally in Section 8 we summarize our findings.

## 2. Stochastic Taylor Series

The main technical tool we will use are stochastic Taylor series. These are described in detail in [24]. In this section we will recall the main definitions and results. We will make some minor notational changes so that we can use the Einstein summation convention.

Let  $X_t$  satisfy a  $d$ -dimensional stochastic differential equation driven by  $m$  independent Brownian motions  $W_t^\alpha$ ,  $\alpha = 1, 2, \dots, m$ . We write

$$dX_t = a(X, t)dt + b_\alpha(X, t)dW_t^\alpha \quad (2.1)$$

where  $X_t$  is an random process taking values on  $\mathbb{R}^d$ .  $a$  and  $b_\alpha$  are also  $\mathbb{R}^d$ -valued for each  $\alpha$ . We are using the Einstein summation convention, which dictates that when there are matching indices in an expression one should take the sum over the given index. Thus (2.1) is an abbreviation for:

$$dX_t = a(X, t)dt + \sum_{\alpha=1}^m b_\alpha(X, t)dW_t^\alpha.$$

The advantage of the Einstein summation convention is not simply that it makes formulae shorter. The convention also makes it easier to spot incorrect formulae. This is because, in formulae that are valid in all coordinate systems, the summed indices should always consist of one upper and one lower index.

In this section we will use Greek indices to index the different Brownian motions and Roman indices to index components of vectors in  $\mathbb{R}^d$ . This additional convention is not strictly necessary as the range of the index can be deduced from the position of the index alone.

A multi-index  $\xi$  is defined to be a finite list of integer numbers between 0 and  $m$  and this definition includes the empty list  $()$ . Let  $l(\xi)$  denote the length of  $\xi$ . Let  $n(\xi)$  denote the number of zeros in  $\xi$ . For  $\xi$  with length greater than 0, we define:  $-\xi$  to be the result of removing the first element from  $\xi$ ;  $\xi-$  for the result of removing the last element;  $\xi_1$  for the first element; and  $\xi_{-1}$  for the last element.

Multi-indices enumerate stochastic integrals with respect to the Brownian motions  $W_t^\alpha$  and time. The following definitions are related to those in [24, p.169]. We define  $W_t^0 := t$  so that the indices equal to 0 correspond to time. We define the multi-integral associated with  $\xi$  by:

$$I_{t_1, t_2}^\xi(f) = \begin{cases} f(t_2) & \text{if } l(\xi) = 0 \\ \int_{t_1}^{t_2} I_{t_1, s}^{\xi-} dW_s^{\xi_{-1}} & \text{otherwise.} \end{cases}$$

For example the multi-index  $(0, 1, 2)$  is associated with integrating with respect first to time, then  $W^1$ , then  $W^2$ .

$$I_{t_1, t_2}^{(0, 1, 2)}(f) = \int_{t_1}^{t_2} \int_{t_1}^u \int_{t_1}^v f(w) dw dW_v^1 dW_u^2.$$

We re-express the notation in [24, p.177, Eqs. 3.1–3.3] by defining differential operators  $L_\xi$  associated to a multi-index as:

$$L_\xi f = \begin{cases} f & \text{if } l(\xi) = 0 \\ \frac{\partial f}{\partial t} + a^i \frac{\partial f}{\partial x^i} + \frac{1}{2} b_\alpha^i b_\beta^j g_E^{\alpha\beta} \frac{\partial^2 f}{\partial x^i \partial x^j} & \text{if } \xi = (0) \\ b_{\xi_1}^i \frac{\partial f}{\partial x^i} & \text{if } l(\xi) = 1 \text{ and } \xi \neq (0) \\ L_{\xi_1}(L_{-\xi} f) & \text{otherwise.} \end{cases}$$

Here  $g_E^{\alpha\beta}$  denotes the covariance matrix of the  $d$  Brownian motions  $W_t^\alpha$ . Since we have assumed that the Brownian motions are independent, this will equal the identity matrix. We choose to write  $g_E$  instead of using the Kronecker delta because it transforms as a tensor of type  $(2, 0)$ . In addition, one can simply replace  $g_E$  with the quadratic co-variation tensor if one wishes to consider SDEs driven by more general continuous semi-martingales.

Since  $L_\xi$  contains a total of  $l(\xi) + n(\xi)$  derivatives,  $L_\xi$  acts on functions in  $C^{l(\xi)+n(\xi)}(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R})$ .

The following definition is related to Eq. (9.1) page 206 in [24].

DEFINITION 1. The *Itô–Taylor expansion* of order  $\gamma = 0, \frac{1}{2}, 1, \dots$  is given by

$$X_t^\gamma = \sum_{l(\xi)+n(\xi) \leq 2\gamma} L_\xi(x)|_{(t_0, X_{t_0})} I_{t_0, t}^\xi(1)$$

where  $x$  denotes the function  $x(t, X) = X$ . When we speak of the expansion of a given order, we will assume that all the necessary derivatives exist.

The Itô–Taylor expansion allows one to approximate  $X_t$  using  $X_t^\gamma$ . Loosely speaking, this approximation will be accurate in mean squared up to order  $\gamma$ . A precise statement is given in Proposition 2.2.

DEFINITION 2. The *weak Itô–Taylor expansion* of order  $\beta = 0, 1, 2, \dots$  is given by

$$\eta_\beta(t) = \sum_{l(\xi) \leq \beta} L_\xi(x)|_{(t_0, X_{t_0})} I_{t_0, t}^\xi(1)$$

where  $x$  denotes the function  $x(t, X) = X$ . When we speak of the expansion of a given order, we will assume that all the necessary derivatives exist.

The weak Itô–Taylor expansion is of interest if one measures the error using the size of the expectation of the error, rather than the expectation of the size of the error. We will give a precise statement in Proposition 2.3.

Given a smooth vector valued function  $f$  defined on  $\mathbb{R}^d$  we have by Itô’s lemma that

$$df(X_t) = L_0(f)dt + L_\alpha(f)dW_t^\alpha. \quad (2.2)$$

The system of equations (2.1) and (2.2) define a higher dimensional SDE. We can use this to compute Itô–Taylor expansions for this higher dimensional system and hence compute approximations to  $f(X_t)$ . This calculation gives rise to the following more general definition.

DEFINITION 3. The *Itô–Taylor expansion* of order  $\gamma = 0, \frac{1}{2}, 1, \dots$  for  $f(X_t)$  is given by

$$f_t^\gamma = \sum_{l(\xi)+n(\xi) \leq 2\gamma} L_\xi(f)|_{(t_0, X_{t_0})} I_{t_0, t}^\xi(1)$$

When we speak of the expansion of a given order, we will assume that all the necessary derivatives exist. The weak Itô–Taylor expansion for  $f(X_t)$  is defined similarly.

By taking expectations of either of these Itô–Taylor expansion we obtain an expansion for  $E(f(X_t))$ . This expansion is not explicitly considered in [24].

DEFINITION 4. The *Itô–Taylor expansion* of order  $\gamma = 0, 1, 2, \dots$  for  $E(f(X_t))$  is given by

$$\overline{f}_t^\gamma = \sum_{k=0}^{\gamma} (L_0)^k f \frac{(t - t_0)^k}{k!}$$

Since the Itô–Taylor expansion for  $E(f(X_t))$  is defined in terms of the coefficients of an SDE it is conceptually distinct from the classical Taylor expansion for  $E(f(X_t))$ . However, it follows from the convergence results we state below that, under suitable assumptions on the coefficients of the SDE, these two series will coincide.

We will wish to compute  $L^2$  norms of terms in our Taylor series. The next result is well-known and provides the basis for the Wiener-Chaos expansion

LEMMA 2.1. We suppose that for all  $i$ ,  $W_0^i = 0$ . Given a time  $t$ , and  $i, j \in \{1, \dots, m\}$ , the integrals

$$\begin{aligned} I_{0,t}^{(0)}(1) &= t \\ I_{0,t}^{(i)}(1) &= W_t^i \\ I_{0,t}^{(i,j)}(1) &= \int_0^t W_s^i dW_s^j \end{aligned}$$

are orthogonal in expectation.

We wish to state some results on the convergence of Itô–Taylor series. We will first need a few more definitions.

First we define spaces  $\mathcal{H}_\xi$  associated with multi-indices  $\xi$ . Associated to the empty index  $()$  we have the set  $\mathcal{H}_{()}$  of adapted cadlag processes  $f_t$  with

$$|f(t, \omega)| < \infty$$

with probability one for each  $t \geq 0$ .  $\mathcal{H}_{(0)}$  consists of the adapted cadlag processes with

$$\int_0^t |f(s, \omega)| ds < \infty$$

with probability one for each  $t \geq 0$ .  $\mathcal{H}_{(\alpha)}$  has the same definition for any positive  $\alpha$ : it is the set of adapted cadlag processes with

$$\int_0^t |f(s, \omega)|^2 ds < \infty$$

with probability one for each  $t \geq 0$ . We now recursively define  $\mathcal{H}_\xi$  for  $\xi$  of length greater than 1 to be the set of adapted cadlag processes such that the integral process  $I_{0,t}^{\xi-}(f)$ , when viewed as a function of  $t$ , lies in  $\mathcal{H}_{\xi_1}$ .

We define  $\mathcal{M}$  to be the set of all multi-indices.

Given a subset  $\mathcal{A} \subseteq \mathcal{M}$  we define the *remainder set*  $\mathcal{B}(\mathcal{A})$  to be the set

$$\mathcal{B}(\mathcal{A}) = \{\xi \in \mathcal{M} \setminus \mathcal{A} : -\xi \in \mathcal{A}\}$$

Thus the remainder set contains all the indices immediately following the indices in  $\mathcal{A}$ . By estimating integrals in the remainder set, one can bound the error of the Itô–Taylor series as we will see below.

We define

$$\Lambda_k = \{\xi \in \mathcal{M} : l(\xi) + n(\xi) \leq k\}.$$

Thus the order  $\gamma$  Itô–Taylor series is a sum over multi-indexes in  $\Lambda_{2\gamma}$ .

We can now state a result on the convergence of the Itô–Taylor series. The following result is a simplified version of Proposition 5.9.1 in [24].

**PROPOSITION 2.2.** *Suppose that  $L_\xi x|_{X_{t_0}, t_0} \in \mathcal{H}_\xi$  for all  $\xi \in \Lambda_k$ . Suppose that  $L_\xi x|_{X_{t,t}} \in \mathcal{H}_\xi$  with*

$$\sup_{0 \leq t \leq T} E \left( |(L_\xi x|_{X_{t,t}})|^2 \right) \leq C_1$$

for all  $\xi \in \mathcal{B}(\Lambda_k)$  and some constant  $C_1$ . Then

$$E \left( |X_t - X_t^{\frac{k}{2}}|^2 \right) \leq C_2 (t - t_0)^{k+1}$$

for some constant  $C_2$ . Here  $X_t^{\frac{k}{2}}$  is the order  $\frac{k}{2}$  Itô–Taylor expansion with  $k = 0, 1, \dots$

This next result on the convergence of weak Itô–Taylor series is a restatement of Proposition 5.11.1 in [24].

**PROPOSITION 2.3.** *Let  $\beta \in \{1, 2, \dots\}$  and  $T \in (0, \infty)$  be given. Let  $C_P^l(\mathbb{R}^d, \mathbb{R})$  denote the space of  $l$  times continuously differentiable functions whose derivatives of order up to and including  $l$  have polynomial growth. Suppose that the components  $a^k$  and  $b^{kj}$  of  $a$  and  $b^j$  are time-independent and satisfy Lipschitz conditions, linear growth bounds and belong to  $C_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ . Then for each  $g \in C_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$  there exist constants  $K \in (0, \infty)$  and  $r \in \{1, 2, \dots\}$  such that*

$$\sup_{0 \leq t \leq T} |E(g(X_t) - g(\eta_\beta(t)))| \leq K (1 + |X_0|^{2r}) T^{\beta+1}$$

where  $\eta_\beta(t)$  is the weak Itô–Taylor series and the expectation is taken conditional on the information at time 0.

In [24], the authors state that they prove their version of our Proposition 2.3 in the case of autonomous SDEs for the sake of simplicity of notation. One expects that a similar result can be proved for time dependent coefficients, but doing so would be a distraction from the main purpose of this paper. As a result, we will occasionally restrict ourselves to considering the case of autonomous SDEs. Where possible we will prefer to use Proposition 2.2 to prove results in fuller generality.

### 3. The jet formulation of SDEs on manifolds

We will be interested in approximating SDEs in  $\mathbb{R}^r$  with SDEs on submanifolds of  $\mathbb{R}^r$ . Thus we need to consider how to formulate SDEs on manifolds. One option is to work extrinsically and consider only manifolds embedded in  $\mathbb{R}^r$ , and to define SDEs on the manifold by considering SDEs on the ambient space whose solutions are confined to the manifold. Another approach is to use an interpretation of SDEs on manifolds where the existence of solutions is defined



intrinsically, but in a coordinate free manner. A survey of various approaches to understanding SDEs on manifolds is given in [4]. In this paper, our presentation will use coordinate charts, so it is natural for us to pursue a coordinate free approach to the existence and uniqueness of solutions to SDEs on manifolds. We will follow the approach taken in [3] of using jets and the notion of convergence in mean square on compacts. We will briefly review this approach.

Let  $M$  be a manifold and  $g$  be a Riemannian metric on  $M$ . Let  $K$  be a compact subset of  $M$ . Let  $d^g$  denote the Riemannian distance function. Let  $K^0$  denote the interior of  $K$ . We define an equivalence relation  $\sim$  on  $M$  by  $x \sim y$  if either  $x = y$  or both  $x \notin K^0$  and  $y \notin K^0$ . The quotient space  $M/\sim$  is simply the one-point compactification of  $K^0$ . We write  $\infty$  for the equivalence class consisting of all points outside  $K^0$ . We may define a semimetric  $\tilde{d}^{g,K}$  on  $M/\sim$  by

$$\tilde{d}^{g,K}([x], [y]) = \inf_{X \sim x, Y \sim y} d^g(X, Y). \quad (3.1)$$

This is not a metric since  $\tilde{d}^{g,K}$  does not obey the triangle inequality. Nevertheless, convergence of a sequence in  $\tilde{d}^{g,K}$  implies convergence in  $M/\sim$ .

Given a stochastic process  $X : [0, T] \rightarrow M \cup \{\infty\}$  and a compact subset  $K$  of  $M$  we define a new stochastic process  $X^K$  by

$$X_t^K(\omega) = \begin{cases} X_t(\omega) & \text{if } X_{t'}(\omega) \in K^0 \text{ for all } t' < t \\ \infty & \text{otherwise.} \end{cases}$$

**DEFINITION 5.** Let  $X^i$  be a sequence of stochastic processes in  $M \cup \{\infty\}$ . For a fixed time  $t$ , we say that  $X^i$  converges to  $X$  in mean square on compacts if for all compact sets  $K \subseteq M$  and Riemannian metrics  $g$  on  $M$

$$E(\tilde{d}^{g,K}((X^i)_t^K, X_t^K)^2) \rightarrow 0 \text{ as } i \rightarrow \infty.$$

Note that this definition of convergence is designed to be manifestly coordinate invariant, but in fact convergence in  $\tilde{d}^{g,K}$  is independent of the choice of Riemannian metric  $g$ . Note also that on a compact manifold convergence in mean square on compacts is equivalent to ordinary convergence in mean square.

In [3], this notion of convergence was combined with the language of jets to show how one can define a notion of an SDE on a manifold.

We will say that an SDE on a manifold  $M$  driven by  $m$  dimensional Brownian motion  $W_t$  is defined by choosing at each point  $x \in M$  and time  $t$  a smooth map

$$\gamma_{x,t} : \mathbb{R}^m \rightarrow M, \quad \gamma(0) = x.$$

For a given time interval  $\delta t$  and starting point  $X_0$ , we may then define a process,  $X^{\delta t}$ , by requiring that at times  $(\delta t, 2\delta t, 3\delta t, \dots)$  we have

$$X_{t+\delta t}^{\delta t} = \gamma_{X_t^{\delta t}, t}(W_{t+\delta t} - W_t). \quad (3.2)$$

We define  $X_{t+\epsilon}^{\delta t} = X_t^{\delta t}$  if  $0 \leq \epsilon \leq \delta t$ . In the case of autonomous SDEs (where  $\gamma_{x,t}$  is independent of  $t$ , it is shown in [3] that  $X^{\delta t}$  converges in mean square on compacts to a unique process  $X$  so long as the choice of  $\gamma_x$  at each point is made smoothly. Moreover, the limiting process depends only on the two jet of  $\gamma_x$  at each point and coincides with the solution to the SDE given in a local chart by

$$dX_t^i = \frac{1}{2} g_E^{\alpha\beta} \frac{\partial^2 \gamma_{X_t}^i}{\partial x^\alpha \partial x^\beta}(0) dt + \frac{\partial \gamma_{X_t}^i}{\partial x^\alpha}(0) dW_t^\alpha.$$

whenever this obeys the classical Lipschitz and growth bounds required to ensure a solution exists.

Conversely, given an SDE of the form (2.1) on  $\mathbb{R}^d$  we define

$$\gamma_{x,t}(V) = b_\alpha(X, t)V^\alpha + \frac{1}{m}g_{\alpha\beta}^E V^\alpha V^\beta \quad (3.3)$$

where  $V^\alpha$  are the components of the vector  $V$  and  $g_{\alpha\beta}^E$  is the Euclidean metric on  $\mathbb{R}^m$  (hence is equal as a matrix to the identity). The results of [3] were given in the case of autonomous SDEs, but the generalization is straightforward.

The SDE (2.1) can be interpreted classically as an equation relating Itô integrals or it can be interpreted in terms of the limiting behaviour of the scheme (3.2). We refer to the latter interpretation as the jet interpretation. The advantage of the jet interpretation of SDEs is that we can find a solution to (2.1) in the jet interpretation whenever the coefficients of (2.1) are smooth. Note that the solution defined by the jet interpretation may blow up to infinity.

Since convergence is defined using expectation in mean square on compacts it is natural to also measure the divergence of stochastic processes using this same measure. The next proposition shows that we can simplify the technical conditions required in Proposition 2.2 if we use this measure.

**PROPOSITION 3.1.** *Let  $X_t$  and  $Y_t$  be processes defined by SDEs of the form (2.1) with smooth coefficients and deterministic initial condition  $X_0 = Y_0$ . Let  $d^K = \tilde{d}^{g,K}$  where  $g$  is the Euclidean metric. Let  $K \subseteq \mathbb{R}^d$  be a compact set containing the ball of radius  $R$  around  $X_0$ . Then*

$$E(d^K(X_t, Y_t)^2) = E(|X_t^{\frac{k}{2}} - Y_t^{\frac{k}{2}}|^2) + O(t^{k+1})$$

where  $X_t^{\frac{k}{2}}$  and  $Y_t^{\frac{k}{2}}$  are the order  $\frac{k}{2}$  Itô-Taylor series for  $X_t$  and  $Y_t$  respectively. Note that the expectation on the right of can be expanded in terms of the Itô-Taylor coefficients of  $X_t$  and  $Y_t$ .

*Proof.* Let  $e_t$  denote the event that either  $X_t$  or  $Y_t$  exits the ball of radius  $\frac{R}{3}$  around  $X_0$  in a time less than or equal to  $t$ . To proceed, we need the martingale estimate [30, Theorem 37.8, p.77], which we state below for the reader's convenience: let  $L$  be a real-valued local martingale such that for all  $t \geq 0$  there exists a deterministic constant  $c_t$  with the property that  $[L]_t \leq c_t$  a.s. Then, for all  $t \geq 0$  and all  $y > 0$

$$P(\max_{0 \leq s \leq t} L_s > y) \leq \exp(-y^2/2c_t) \quad (3.4)$$

Let

$$\tau := \min\{t \geq 0 : |(X_t, Y_t) - (X_0, X_0)| \geq R/3\} \quad (3.5)$$

The Itô formula yields the a decomposition  $|(X_t, Y_t) - (X_0, X_0)|^2 = L_t + A_t$  with  $L_t$  sum of Brownian integrals and  $A_t$  time integral, all of which for  $t \leq \tau \wedge \varepsilon$  (any  $\varepsilon > 0$ ) have bounded integrand, (this is by continuity of the coefficients and compactness of  $\overline{B}_{R/3}(X_0) \times \overline{B}_{R/3}(X_0) \times [0, \varepsilon]$ ). By the Kunita-Watanabe identity, also  $[L]_t$  can be expressed as a time integral with bounded integrand: let  $S > 0$  bound the sum of the absolute values of all integrands mentioned for  $t \in [0, \tau \wedge \varepsilon]$ . Then, still on  $t \leq \tau \wedge \varepsilon$  we have  $|A_t|, [L]_t \leq St$  and

thus  $|(X_t, Y_t) - (X_0, X_0)|^2 \leq L_t + S\varepsilon$ . Picking  $\varepsilon = (R/3)^2/(3S)$  we have, for  $0 \leq t \leq \varepsilon$

$$\begin{aligned} P(e_t) &\leq P\left(\max_{0 \leq s \leq t} |(X_s, Y_s) - (X_0, X_0)|^2 \geq (R/3)^2\right) \\ &= P\left(\max_{0 \leq s \leq \tau \wedge t} |(X_s, Y_s) - (X_0, X_0)|^2 \geq (R/3)^2\right) \\ &\leq P\left(\max_{0 \leq s \leq \tau \wedge t} L_s > (R/3)^2/2\right) \\ &= P\left(\max_{0 \leq s \leq t} L_{\tau \wedge s} > (R/3)^2/2\right) \\ &\leq \exp\left(-\frac{C}{t}\right) \end{aligned} \tag{3.6}$$

for some  $0 < C = C(R, S)$ , where the last inequality is an application of (3.4) to the stopped martingale  $L_{\tau \wedge \cdot}$ . This implies  $P(e_t) = O(t^k)$  for any  $k$ .

$$\begin{aligned} E(d^K(X_t, Y_t)^2) &= E(d^K(X_t, Y_t)^2 \mid \neg e_t)P(\neg e_t) + E(d^K(X_t, Y_t)^2 \mid e_t)P(e_t) \\ &= E(d^K(X_t, Y_t)^2 \mid \neg e_t)P(\neg e_t) + O(t^{k+1}). \end{aligned}$$

Here we have used the fact that  $E(d^K(X_t, Y_t)^2 \mid e_t)$  can be bounded above (since  $K$  is compact), together with (3.6). If  $X_t, Y_t \in B_{\frac{R}{3}}(X_0)$  and  $B_R(X_0) \subseteq K$  we have that  $d^K(X_t, Y_t)^2 = |X_t - Y_t|^2$ . Hence

$$E(d^K(X_t, Y_t)^2) = E(|X_t - Y_t|^2 \mid \neg e_t)P(\neg e_t) + O(t^{k+1}). \tag{3.7}$$

Let us take the SDEs defining the processes  $X_t$  and  $Y_t$  and modify the coefficients so that they remain fixed inside  $B_{\frac{R}{3}}(X_0)$  but smoothly drop to zero to ensure that they are compactly supported in  $B_r(X_0)$ . Let us write  $\tilde{X}_t$  and  $\tilde{Y}_t$  for the solutions of the modified SDEs. By construction

$$E(|X_t - Y_t|^2 \mid \neg e_t) = E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid \neg e_t). \tag{3.8}$$

We have

$$E(|\tilde{X}_t - \tilde{Y}_t|^2) = E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid \neg e_t)P(\neg e_t) + E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid e_t)P(e_t).$$

We rearrange this to find

$$\begin{aligned} E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid \neg e_t) &= \frac{1}{P(\neg e_t)}E(|\tilde{X}_t - \tilde{Y}_t|^2) - \frac{P(e_t)}{P(\neg e_t)}(E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid e_t)) \\ &= E(|\tilde{X}_t - \tilde{Y}_t|^2) + \frac{P(e_t)}{1 - P(e_t)}(E(|\tilde{X}_t - \tilde{Y}_t|^2) - E(|\tilde{X}_t - \tilde{Y}_t|^2 \mid e_t)) \\ &= E(|\tilde{X}_t - \tilde{Y}_t|^2) + O(t^{k+1}). \end{aligned} \tag{3.9}$$

To obtain the last line, we have used the fact that by  $|\tilde{X}_t - \tilde{Y}_t|$  is bounded because the dynamics of  $X_t$  and  $Y_t$  restrict them to lying in  $B_r(X_0)$  together with the bound (3.6).

Next we observe that by Proposition 2.2

$$\begin{aligned} E(|\tilde{X}_t - \tilde{Y}_t|^2) &= E(|\tilde{X}_t^{\frac{k}{2}} - \tilde{Y}_t^{\frac{k}{2}}|^2 + O(t^{k+1})) \\ &= E(|X_t^{\frac{k}{2}} - Y_t^{\frac{k}{2}}|^2 + O(t^{k+1})) \end{aligned} \tag{3.10}$$

since the Itô–Taylor series for  $\tilde{X}$  and  $\tilde{Y}$  coincide with the Itô–Taylor series for  $X$  and  $Y$ . Combining (3.7), (3.8), (3.9) and (3.10) yields the desired result.  $\square$

Our next proposition uses a similar device can be used to simplify the technical conditions of Proposition 2.3. The proof is essentially identical to that of Proposition 3.1 and so is omitted.

PROPOSITION 3.2. Let  $X_t$  and  $Y_t$  be processes on  $\mathbb{R}^{d_1}$  and  $\mathbb{R}^{d_2}$  respectively which are defined by SDEs of the form (2.1) with smooth, time-independent coefficients and deterministic initial condition  $X_0 = Y_0$ . Let  $f : (\mathbb{R}^{d_1} \cup \{\infty\}) \times (\mathbb{R}^{d_2} \cup \{\infty\}) \rightarrow \mathbb{R}$  be bounded and smooth in some neighbourhood of  $(X_0, Y_0)$ , then

$$E(f(X_t, Y_t)) = E(\bar{f}^k) + O(t^{k+1})$$

where  $\bar{f}^k$  is the order  $k$  Itô-Taylor series for  $E(f(X_t, Y_t))$ .

Note that in the case of SDEs with time-independent coefficients, Proposition 3.1 follows from Proposition 3.2. Hence if the reader accepts that Proposition 2.3 can be generalized to the case of time varying coefficients, Proposition 3.1 is redundant.

#### 4. Projecting stochastic differential equations

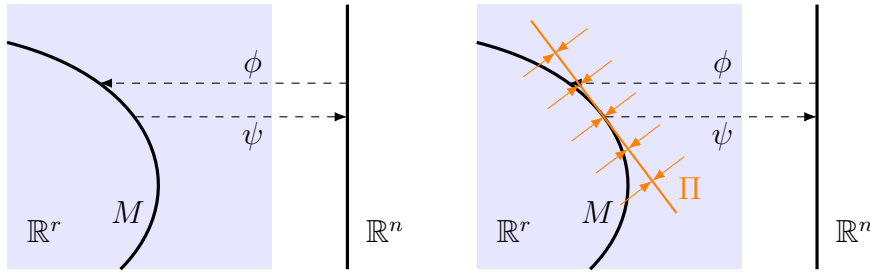


FIGURE 2. Left: An  $n$ -dimensional manifold  $M$  in  $\mathbb{R}^r$ ,  $r > n$ . Right: Tangent space linear projection used in the Stratonovich and Itô-vector projections

Let  $M$  be an  $n$ -dimensional submanifold of  $\mathbb{R}^r$  with chart  $\psi : U \rightarrow \mathbb{R}^n$  for some open neighbourhood  $U$  in  $M$ . The inverse  $\phi = \psi^{-1}$  gives an embedding of  $\text{Im } \psi$  into  $\mathbb{R}^r$ . The setup is illustrated in Figure 2.

Suppose we are given an Itô SDE on  $\mathbb{R}^r$ ,  $dX_t = a(X_t, t) dt + b_\alpha(X_t, t) dW_t^\alpha$ , that we write in concise form as

$$dX = a dt + b_\alpha dW_t^\alpha, \quad X_0 \quad (4.1)$$

with  $X_0 \in M$ .

We wish to find an SDE on  $\mathbb{R}^n$  of the form  $dY_t = A(Y_t, t) dt + B_\alpha(Y_t, t) dW_t^\alpha$ , again written concisely as

$$dY = A dt + B_\alpha dW_t^\alpha, \quad Y_0 = \psi^{-1}(X_0), \quad (4.2)$$

whose mapped solution  $\phi(Y)$  in some sense approximates the solution  $X$  of the original equation on  $\mathbb{R}^r$ . We will consider three approaches.

##### 4.1. Stratonovich Projection

DEFINITION 6. Let  $W_t$  be an  $\mathbb{R}^m$  valued Brownian motion. Given a Stratonovich SDE on  $\mathbb{R}^r$

$$dX = \bar{a} dt + b_\alpha \circ dW_t^\alpha$$

and a chart  $\psi : U \rightarrow \mathbb{R}^n$  for some neighbourhood in  $M$  we define the *Stratonovich projection* of the SDE to be:

$$dY = \bar{A} dt + B_\alpha \circ dW_t^\alpha$$

where:

$$\bar{A}(Y_t, t) = (\psi_*) \Pi_{\phi(Y_t)}(\bar{a}(\phi(Y_t), t)) \quad (4.3)$$

$$B_\alpha(Y_t, t) = (\psi_*) \Pi_{\phi(Y_t)}(b_\alpha(\phi(Y_t), t)) \quad (4.4)$$

where  $\Pi$  is the projection of  $\mathbb{R}^r$  onto  $\phi_*(\mathbb{R}^n)$  defined by the Euclidean metric.

Because we know that projection of vector fields can be defined similarly, and because we know that the coefficients of Stratonovich SDEs transform like vector fields, we see that the definition above defines a Stratonovich SDE on  $M$ . Indeed, if one is willing to accept that projection of vector fields onto a submanifold is well-defined, then one could define the projection of a Stratonovich SDE as the projection of the coefficient functions.

Trying the same method for an Itô SDE does not work. One cannot simply apply projection to the coefficient functions of an Itô SDE because the coefficients of an Itô SDE on a manifold do not transform like vector fields.

The *Stratonovich projection* of an Itô SDE is trivially defined by the recipe:

- (i) rewrite the Itô SDE as a Stratonovich SDE;
- (ii) apply the Stratonovich projection as defined above;
- (iii) rewrite the resulting Stratonovich SDE as an Itô SDE.

In other words, while the definition of Stratonovich projection is most conveniently expressed using Stratonovich calculus, the notion of projection is independent of the calculus used to write down the differential equations.

Linear projection provides the best possible way to approximate vectors in  $\mathbb{R}^r$  with vectors in  $T_X M$ . For ODEs, this implies that the projected ODE is the best possible approximation in  $M$  of the original ODE. However, the situation is different for SDEs. It is not immediately clear how good an approximation the projected Stratonovich SDE solution  $\phi(Y)$  is for the original SDE  $X$  solution. For example, we cannot immediately extend the optimality argument for ODEs to Stratonovich SDEs pathwise, because of the rough paths property of SDEs solutions. In this sense, with the information we have given so far, the definition of the Stratonovich projection is motivated by purely heuristic considerations. Nevertheless, the Stratonovich projection gives good results when applied to approximation of non-linear filtering problems (see [10], [11], [5]) and we will discuss optimality arguments later on, when discussing the Itô-vector projection, and illustrate the time-symmetric optimality of the Stratonovich projection more in detail.

In the next sections we will use optimality arguments to derive two alternative notions of projection.

#### 4.2. Itô-vector projection

We wish to consider the minimization problem of finding coefficients  $A$  and  $B$  such that the solution of the SDE (4.2) has the property that  $\phi(Y_t)$  is, in some sense, as close to the solution  $X_t$  of (4.1) as possible.

The next proposition shows how to give a precise meaning to this notion using the Itô–Taylor expansion.

PROPOSITION 4.1. *Let  $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^d$  and  $F : \mathbb{R}^{d_y} \rightarrow \mathbb{R}^d$  be smooth maps. Let  $x$  be a process on  $\mathbb{R}^{d_x}$  and  $y$  be a process on  $\mathbb{R}^{d_y}$  given by:*

$$\begin{aligned} dx_t &= a(x_t, t) dt + b_\alpha(x_t, t) dW_t^\alpha, & x_0 \\ dy_t &= A(y_t, t) dt + B_\alpha(y_t, t) dW_t^\alpha, & y_0 \end{aligned} \quad (4.5)$$

with  $f(x_0) = F(y_0)$ . Define

$$z_t = f(x_t) - F(y_t).$$

Let  $z_t^i$  denote the components of the order  $i$  Itô–Taylor expansion for  $z$ . We have that:

$$\begin{aligned} E(|z_t^{\frac{1}{2}}|^2) &= \sum_{\alpha} |f_*(b_\alpha(x_0, 0)) - F_*(B_\alpha(y_0, 0))|^2 t \\ E(|z_t^1|^2) &= \sum_{\alpha} |f_*(b_\alpha(x_0, 0)) - F_*(B_\alpha(y_0, 0))|^2 t \\ &\quad + \left( \left| f_*(a(x_0, 0)) - F_*(A(y_0, 0)) \right. \right. \\ &\quad + \frac{1}{2} (\nabla_{b_\alpha(x_0, 0)} f_*) b_\beta(x_0, 0) g_E^{\alpha\beta} - \frac{1}{2} (\nabla_{B_\alpha(y_0, 0)} F_*) B_\beta(y_0, 0) g_E^{\alpha\beta} \Big|^2 \\ &\quad \left. \left. + \mathcal{R}(f, F, b, B)^2 \right) t^2 \end{aligned} \quad (4.6)$$

where  $\mathcal{R}(f, F, b, B)$  is a term independent of  $a$ ,  $A$  and  $t$ .

*Proof.* As an example of how to compute the operators  $L_\xi$  for the system of equations (4.5), we write down  $L_{(\alpha)}$ .

$$L_{(\alpha)} f = b_\alpha^i \frac{\partial f}{\partial x^i} + B_\alpha^i \frac{\partial f}{\partial y^i}.$$

Let us now the first few terms of the Itô–Taylor expansion for  $z = f(x) - F(y)$ .

$$\begin{aligned} L_{(0)}(z) &= f_*(a(x_t, t)) - F_*(A(y_t, t)) \\ &\quad + \frac{1}{2} (\nabla_{b_\alpha(x_t, t)} f_*) b_\beta(x_t, t) g_E^{\alpha\beta} - \frac{1}{2} (\nabla_{B_\alpha(y_t, t)} F_*) B_\beta(y_t, t) g_E^{\alpha\beta}. \\ L_{(\alpha)}(z) &= f_*(b_\alpha(x_t, t)) - F_*(B_\alpha(y_t, t)). \\ L_{(\alpha, \beta)}(z) &= L_{(\alpha)} L_{(\beta)}(z) \\ &= b_\alpha^i(x_t, t) \frac{\partial}{\partial x^i} f_*(b_\beta(x_t, t)) - B_\alpha^i(y_t, t) \frac{\partial}{\partial y^i} F_*(B_\beta(y_t, t)). \end{aligned}$$

We can now write down the order 1 Itô–Taylor expansion  $z_t^1$ . It is

$$\begin{aligned} z_t^1 &= \left( f_*(a(x_0, 0)) - F_*(A(y_0, 0)) \right. \\ &\quad + \frac{1}{2} (\nabla_{b_\alpha(x_0, 0)} f_*) b_\beta(x_0, 0) g_E^{\alpha\beta} - \frac{1}{2} (\nabla_{B_\alpha(y_0, 0)} F_*) B_\beta(y_0, 0) g_E^{\alpha\beta} \Big) I_{0,t}^{(0)} \\ &\quad + \left( f_*(b_\alpha(x_0, 0)) - F_*(B_\alpha(y_0, 0)) \right) I_{0,t}^{(\alpha)} \\ &\quad + \left( b_\alpha^i(x_0, 0) \frac{\partial}{\partial x^i} f_*(b_\beta(x_0, 0)) - B_\alpha^i(y_0, 0) \frac{\partial}{\partial y^i} F_*(B_\beta(y_0, 0)) \right) I_{0,t}^{(\alpha, \beta)} \end{aligned}$$

We can now use Lemma 2.1 to calculate  $E(|z_t^1|^2)$ . This gives the desired result.  $\square$

REMARK 1. For readers familiar with the traditional Itô formula in Euclidean spaces, the term  $(\nabla_{b_\alpha(x_0, 0)} f_{i,*}) b_\beta(x_0, 0) g_E^{\alpha\beta}$  for the  $i$ -th component of  $f$  might be more familiar when

written as

$$(\nabla_{b_\alpha f_{i,*}})b_\beta g_E^{\alpha\beta} = \text{Tr} [b^T(Hf_i)b]$$

where  $\text{Tr}$  is the trace operator and  $H$  is the Hessian operator.

We will now attempt to choose coefficients for the equation (4.2) on the manifold so that its solution is as close as possible to the solution of (4.1) in Euclidean space. As a first step to doing this we consider how to minimize the terms in the Itô–Taylor expansion of the difference  $X_t - \phi(Y_t)$ .

**THEOREM 4.2** (Itô–Taylor series and Itô-vector projection). *If we wish to find smooth coefficients  $A$  and  $B$  for equation (4.2) which minimize the mean square ( $L^2$ ) norm of the order  $\frac{1}{2}$  Itô–Taylor series for  $X_t - \phi(Y_t)$  for time  $t_0 = 0$ , we must take*

$$B_\alpha(Y_0, 0) = (\psi_*)_{X_0} \Pi_{X_0} b_\alpha(X_0, 0) \quad (4.7)$$

where  $\Pi_{X_0}$  is the projection map onto the tangent space of  $M$  at  $X_0$ . If we now suppose that  $B$  is chosen so that this minimum is achieved at all points of  $U$ , a neighbourhood of  $X_0$  in  $M$ , then the mean square  $L^2$  norm of the order 1 Itô–Taylor series is minimized by taking

$$A(Y_0, 0) = (\psi_*)_{X_0} \Pi_{X_0} \left( a(X_0, 0) - \frac{1}{2} (\nabla_{B_\alpha(Y_0, 0)} \phi_*) B_\beta(Y_0, 0) g_E^{\alpha\beta} \right).$$

*Proof.* We apply Proposition 4.1 taking  $f$  equal to the identity,  $F$  equal to  $\phi$ ,  $x_t = X_t$  and  $y_t = Y_t$ . To minimize the order  $\frac{1}{2}$  Itô–Taylor series for  $X_t - \phi(Y_t)$  we must solve the problem:

$$\text{Find } B_\alpha(Y_0, 0) \text{ minimizing } \sum_{\alpha} |\phi_*(B_\alpha(Y_0, 0)) - b_\alpha(X_0, 0)|^2.$$

The solution to this is given by  $B_\alpha(Y_0, 0) = \psi_*(V_\alpha)$  where the vectors  $V_\alpha$  give a solution to the problem:

$$\text{Find } V_\alpha \in \text{Im } \phi_* \text{ minimizing } \sum_{\alpha} |V - b_\alpha(X_0, 0)|^2.$$

The standard properties of the projection map tell us that  $V_\alpha = \Pi_{X_0} b_\alpha(X_0, 0)$ .

The same argument is used to find the formula for the coefficient  $A$  that minimizes the order 1 Itô–Taylor expansion.  $\square$

**REMARK 2.** The optimal  $B$  in the above definition is the same we had in the Stratonovich projection in Eq. (4.4). The optimal  $A$  is different.

**COROLLARY 4.3.** *Let  $K$  be a compact set in  $\mathbb{R}^r$  containing a neighbourhood of  $X_0$  with  $K \cup M$  closed. Let  $d^K$  be the semimetric defined by (3.1) where we take  $g$  to be the Euclidean metric. If  $Y$  is the process given by (4.2),  $X$  is the process given by (4.1) and we assume both these equations have smooth coefficients then*

$$E(d^K(X_t, \phi(Y_t))^2) = C^B t + O(t^{\frac{3}{2}})$$

for some constant  $C^B$  depending on the coefficient  $B$ .  $C^B$  is minimized by taking  $B$  as defined in (4.7). Note that this choice is independent of  $K$ .

If the coefficients of the SDEs are compactly supported we may replace the  $d_K$  with the Euclidean metric.

We remark that our device of using  $d_K$  and the corresponding notion of convergence in mean square on compacts allows us to avoid imposing technical conditions on the chart  $\psi$ .

*Proof.* Let  $e_t$  be the error term  $X_t - \phi(Y_t)$ . Let  $e_t^{\frac{1}{2}}$  be its order  $\frac{1}{2}$  Ito–Taylor series.

By Proposition 3.1, the expectation  $E(d^K(X_t, \phi(Y_t))^2)$  only depends upon the coefficients of the SDEs in a neighbourhood of  $X_0$ . So we may assume without loss of generality that the coefficients of the SDEs are compactly supported and that  $E(d^K(X_t, Y_t)^2) = E(|e_t|^2)$ .

Since  $e_t^{\frac{1}{2}}$  is an order  $\frac{1}{2}$  Ito–Taylor expansion, we have that  $E(|e_t^{\frac{1}{2}}|) = O(t^{\frac{1}{2}})$ . Our assumption that the coefficients of the SDEs are compactly supported allows us to apply Proposition 2.2 to show  $E(|e_t - e_t^{\frac{1}{2}}|) = O(t)$ .

We now compute

$$E(d^K(X_t, Y_t)^2) = E(|e_t|^2) = E(|e_t^{\frac{1}{2}} + (e_t - e_t^{\frac{1}{2}})|^2) = E(|e_t^{\frac{1}{2}}|^2) + O(t^{\frac{3}{2}}).$$

The result now follows from Theorem 4.2.  $\square$

Theorem 4.2 above motivates the following definition.

**DEFINITION 7.** The Itô-vector projection of the SDE (4.1) onto the manifold  $M$  is given in the chart  $\psi$  by the SDE (4.2) with

$$\begin{aligned} \phi &:= \psi^{-1} \\ B_\alpha(Y_t, t) &:= (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} b_\alpha(\phi(Y_t), t) \\ A(Y_t, t) &:= (\psi_*)_{\phi(Y_t)} \Pi_{\phi(Y_t)} \left( a(\phi(Y_t), t) - \frac{1}{2} (\nabla_{B_\alpha(Y_t, t)} \phi_*) B_\beta(\phi(Y_t), t) g_E^{\alpha\beta} \right) \end{aligned} \quad (4.8)$$

We will demonstrate that the Itô-vector projection is distinct from the Stratonovich projection by calculating an explicit example in Section 7.

One criticism of the above derivation of the Itô-vector projection is that it is peculiar to worry about minimizing a term of order 1 when we cannot even ensure that the projection is accurate to order  $\frac{1}{2}$ . Because of Corollary 4.3, It seems uncontroversial that choosing the diffusion coefficient  $B$  by the prescription above will yield the best approximation, but will it make much difference to choose  $A$  in the same way? In particular, we cannot give an interpretation of our choice of  $A$  in terms of minimizing the mean square distance analogous to Corollary 4.3.

However, choosing  $A$  is important because the errors of order  $\frac{1}{2}$  due to the approximation of  $b$  will cancel on average. The correct choice of  $A$  yields the optimal average value for the approximation. This is made precise by the next result and its corollary.

**THEOREM 4.4** (Itô-vector projection and weak Itô–Taylor expansion). *If we wish to choose the coefficient  $A$  of (4.2) to minimize the norm of the expectation of the order 1 weak Itô–Taylor series for  $X_t - \phi(Y_t)$ , we must take:*

$$A(Y_0, 0) = (\psi_*)_{X_0} \Pi_{X_0} \left( a(X_0, 0) - \frac{1}{2} (\nabla_{B_\alpha(Y_0, 0)} \phi_*) B_\beta(Y_0, 0) g_E^{\alpha\beta} \right). \quad (4.9)$$

where  $\phi = \psi^{-1}$ .

*Proof.* The expectation of the weak Itô Taylor expansion of  $X_t - \phi(Y_t)$  is

$$\left( a(X_0, 0) - \phi_*(A(Y_0, 0)) - \frac{1}{2} (\nabla_{B_\alpha(Y_0, 0)} \phi_*) B_\beta(Y_0, 0) g_E^{\alpha\beta} \right) t.$$



The result now follows immediately from the properties of  $\Pi$ .  $\square$

**COROLLARY 4.5.** *Suppose we are in the situation of Theorem 4.4 except that, for convenience, we will only consider autonomous SDEs. Let  $K$  be a compact set containing a neighbourhood of  $X_0$ . Let  $\mathbf{1}_K$  be its indicator function. Let  $B$  be chosen to minimize  $d^K(X_t, Y_t)$ . Then*

$$E(\mathbf{1}_K(X_t)\mathbf{1}_K(Y_t)(X_t - \phi(Y_t))) = C^A t + O(t^2)$$

for some constant  $C^A$  depending on the coefficient  $A$ .  $|C^A|$  is minimized by taking  $A$  as defined in (4.9). Note that this is independent of the choice of  $K$ . If the coefficients of the SDE are compactly supported, the indicator functions  $\mathbf{1}_K$  are not needed.

*Proof.* Apply Proposition 3.2.  $\square$

**COROLLARY 4.6.** *The Itô-vector projection defines an SDE on the manifold  $M$ . By this we mean that SDE defined on the manifold  $M$  transforms according to Itô's lemma as we change chart  $\psi$ . See [3] for a more detailed discussion of the Itô formulation of SDEs on manifolds.*

*Proof.* The criteria we are using for finding the optimal coefficients of the SDE is given in terms of an estimate of the growth of the difference between the solution to the SDE in  $\mathbb{R}^r$  and the solution to the SDE on the manifold. Since they are expressed in terms of the solutions to the SDE rather than the coefficients of the SDE, the criterion is independent of the choice of chart  $\psi$ .

It follows that the condition we have derived on the coefficients will transform according to Itô's lemma as we change the choice of chart. For an alternative proof by brute-force calculation see [6].  $\square$

We may summarize our results informally as follows: the Itô-vector projection is the choice of  $A$  and  $B$  that simultaneously minimizes the expectation of the error to order  $\frac{1}{2}$  and the error of the expectation to order 1.

Given the Stratonovich–Taylor expansions described in [24], one might wonder if there are versions of Theorems 4.2 and 4.4 using Stratonovich–Taylor series in place of Itô–Taylor series? Might these provide a justification for the Stratonovich projection? The answer is negative. The order 1 Stratonovich–Taylor expansion of [24] is in fact equal to the order 1 Itô–Taylor expansion. The difference is simply that the Stratonovich–Taylor expansion is expressed in terms of Stratonovich coefficients and Stratonovich integrals rather than Itô coefficients and integrals. Thus there is no different “Stratonovich” version of Theorem 4.2.

However, there is a sense in which the Stratonovich projection is optimal in relation with time symmetry. We will address this optimality after introducing two different optimal approximations, the Itô vector and Itô jet projections.

#### 4.3. Itô-jet projection

We now suppose that the open set  $U$  inside our manifold  $M$  has been chosen so that we can find a tubular neighbourhood  $N$  of  $U$  such that the metric projection  $\Pi^s$  is smoothly defined on  $N$ . We then extend  $\Pi^s$  smoothly to the whole of  $\mathbb{R}^n$  in an arbitrary fashion. The metric projection is the map sending a point  $x \in \mathbb{R}^r$  to the nearest point in  $N$ . The standard theory of tubular neighbourhoods tells us that if we choose  $U$  small enough, these conditions will apply.

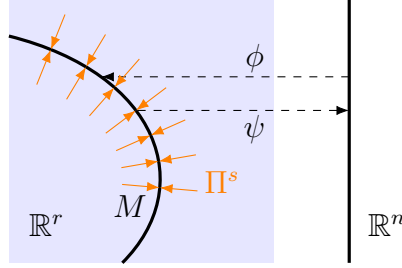


FIGURE 3. Metric projection  $\Pi^s$  of a tubular neighbourhood of  $M$  in  $\mathbb{R}^r$  onto a neighbourhood  $U$  in  $M$ . This is used to define the Itô-jet projection.

Note that the superscript  $s$  in  $\Pi^s$  is short for smooth and is intended to distinguish this map from the linear projection operator  $\Pi_{\phi(x)}$  onto the tangent space at  $\phi(x)$  ( $x \in \mathbb{R}^n$ ). Since the metric on  $M$  is induced by the  $\mathbb{R}^r$  Euclidean metric, we will have that the tangent-space linear projection,  $\Pi_{\phi(x)}$ , will be the first-order-component or best-linear-approximation of the metric projection,  $\Pi^s$ . See also our explicit calculation in Section 6 later on.

For ODEs only the first order linear component of the metric projection is necessary to define the projection. However, Itô SDEs involve explicit second order effects, so that there is an actual difference in applying the tangent vector projection or the full metric projection, going beyond the linear term, in approximating a SDE on a submanifold. As we pointed out in [3], an Itô SDE can be interpreted as a 2-jet. It is then not completely surprising that the second order terms of the metric projection play an important role in understanding the projection of SDEs.

More specifically, in this section we will solve the problem of finding an SDE on the manifold  $M$ ,  $Y_t$  in  $\psi$  coordinates, which minimizes the mean square of the truncated Taylor expansion of the  $M$  geodesic distance between  $\Pi^s(X_t)$  and  $\phi(Y_t)$ , or ambient  $\mathbb{R}^r$  distance between these two points of  $M$ . The two distances will lead to the same result. We call this solution the Itô-jet projection. By contrast, the Itô-vector projection focuses on the  $\mathbb{R}^r$  distance between  $\phi(Y_t)$  and  $X_t$ . Thus the Itô-jet projection uses the metric projection of  $X$  as a benchmark to obtain an optimal approximation  $\phi(Y)$ , whereas the Itô-vector projection uses directly the original  $X$  as a benchmark.

**DEFINITION 8.** Let  $W_t^\alpha$  be independent Brownian motions with  $1 \leq \alpha \leq m$ . Let  $\gamma_x : \mathbb{R}^m \rightarrow \mathbb{R}^r$  be a smoothly varying family of maps satisfying  $\gamma_x(0) = x$  for all  $x \in \mathbb{R}^r$ . We interpret  $\gamma$  as defining an Itô SDE using the ideas of Section 3. We define the Itô-jet projection to be the SDE associated with  $\Pi^s \circ \gamma_y : \mathbb{R}^m \rightarrow M$ .

Since this definition only depends upon germs of  $\Pi^s$  and  $\gamma$ , the Itô-jet projection does not depend upon issues such as the tubular neighbourhood used to define  $\Pi^s$  or the choice of extension of  $\Pi^s$  outside the tubular neighbourhood.

We wish to show that the Itô-jet projection solves the problem of finding the best approximation to the SDE on the manifold, if one measures the quality of the approximation using the truncated Itô-Taylor expansion of either the geodesic distance or the distance in the ambient space  $\mathbb{R}^r$ .

**THEOREM 4.7** (Itô-jet projection as optimal approximation). *Let  $\lambda(x, y)$  denote the square of the geodesic distance between two points on  $M$ . Let  $|x - y|_r^2$  denote the square of the distance between two points in the ambient space. We wish to choose coefficients for equation (4.2) to*

approximate (4.1). The order 1 Itô-Taylor expansion of  $E(\lambda(\Pi^s(X_t), Y_t))$  and of  $E(|\Pi^s(X_t) - Y_t|_r^2)$  vanishes if and only if we take

$$B_\alpha(Y_0, 0) = (\psi_*)_{X_0} \Pi_{X_0} b_\alpha(X_0, 0) \quad (4.10)$$

If we use this to define  $B$  at all points of  $M$ , we have that the order 2 Itô-Taylor expansion of  $E(\lambda(\tilde{\Pi}^s(X_t), Y_t))$  and of  $E(|\tilde{\Pi}^s(X_t) - Y_t|_r^2)$  are both minimized by ensuring that the the 2-jet associated with (4.2) at  $(Y_0, 0)$  is given by  $\Pi^s \circ \gamma_{X_0}$  where  $\gamma_x$  is the 2-jet associated with (4.1). This results in the following drift for (4.2):

$$A(Y_0, 0) = \tilde{\Pi}_*^s(a(X_0, 0)) + \frac{1}{2}(\nabla_{b_{\alpha(X_0, 0)}} \tilde{\Pi}_*^s) b_\beta(X_0, 0) g_E^{\alpha\beta}, \quad (4.11)$$

where we define  $\tilde{\Pi}^s = \psi \circ \Pi^s$ .

*Proof.* We will first prove the result for the geodesic distance.

It will suffice to prove the result in a single chart. Hence we may assume that our coordinates are normal coordinates based at  $X_0$ .

We have the following Taylor series expansion for the square of the geodesic distance (see for example formula 3.4.3 in [9]):

$$\begin{aligned} \lambda(x, y) &= g_{ij}^E(x^i - y^i)(x^j - y^j) \\ &\quad - \frac{1}{12} R_{ikjl}(x^i + y^i)(x^k - y^k)(x^j + y^j)(x^l - y^l) + O((|x| + |y|)^5). \end{aligned} \quad (4.12)$$

The first term is just the Euclidean metric on  $\mathbb{R}^n$ , the term  $R_{ikjl}$  denotes the Riemann curvature tensor of  $M$  at the origin.

The argument of Corollary 4.3 shows that the leading order term of the series for  $E(|\tilde{\Pi}^s(X_t) - Y_t|^2)$  is the  $L^2$  norm of the leading order term of the series for  $\tilde{\Pi}^s(X_t) - Y_t$ . Hence using Proposition 4.1, taking  $f = \tilde{\Pi}^s$  and  $F$  to be the identity, we find that the order 1 Itô-Taylor expansion of  $E(|\tilde{\Pi}^s(X_t) - Y_t|^2)$  is:

$$\sum_{\alpha} |\tilde{\Pi}_*^s(b_\alpha(X_0, 0)) - B_\alpha(Y_0, 0)|^2 t. \quad (4.13)$$

Since the curvature terms of equation (4.12) are all of order 4, we see that they do not contribute to the order 1 Itô-Taylor expansion of  $E(\lambda(\Pi^s(X_t), Y_t))$ . This is because the differential operators  $L_\xi$  in this expansion are all order 2 or less. Hence the order 1 Itô-Taylor expansion of  $E(\lambda(\Pi^s(X_t), Y_t))$  vanishes if and only if equation (4.10) holds.

Assuming this holds, the order  $\frac{1}{2}$  Itô-Taylor series for  $\tilde{\Pi}^s(X_t) - Y_t$  vanishes. So the top order term of the series for  $E(|\tilde{\Pi}^s(X_t) - Y_t|^2)$  is given by the  $L^2$  norm of the order 1 expansion for  $\tilde{\Pi}^s(X_t) - Y_t$ . So, using Proposition 4.1, taking  $f = \tilde{\Pi}^s$  and  $F$  to be the identity, the order 2 Itô-Taylor series for  $E(|\tilde{\Pi}^s(X_t) - Y_t|^2)$  is:

$$\left( \left| \tilde{\Pi}_*^s(a(X_0, 0)) + \frac{1}{2}(\nabla_{b_{\alpha(X_0, 0)}} \tilde{\Pi}_*^s) b_\beta(X_0, 0) g_E^{\alpha\beta} - A(Y_0, 0) \right|^2 + \mathcal{R}(\tilde{\Pi}^s, b, B)^2 \right) t^2 \quad (4.14)$$

where  $\mathcal{R}(\tilde{\Pi}^s, b, B)$  is a term independent of  $a$  and  $A$ .

The non-zero terms in the expectation of the order 2 Itô expansion for  $E(\lambda(\tilde{P}i^2(X, t), Y_t))$  correspond to the multi-indices  $()$ ,  $(0)$  and  $(0, 0)$ . Since the curvature term is fourth order, the only term that will contain a curvature term corresponds to the index  $(0, 0)$ . Moreover, only the highest order term of the operator  $L_{0,0}$  is influenced by the curvature. The coefficient of this highest order term may involve only  $b$  and  $B$  but will not involve  $a$  or  $A$ .

Thus the order 2 Itô-Taylor expansion for  $E(\lambda(\tilde{\Pi}^s(X, t), Y_t))$  is of the form (4.14) since any curvature correction can be absorbed into the term  $\mathcal{R}(\tilde{\Pi}^s, b, B)^2$ . We deduce that the order 2

Itô–Taylor series is minimized by taking  $A$  as in Equation (4.11). When these conditions are rewritten in the language of 2-jets, we get the desired result for the metric  $\lambda$ .

The proof for the metric  $|\cdot|$  follows from Lemma 4.9 given below, and is otherwise essentially identical to that for  $\lambda$ .  $\square$

Note that in this argument we can ensure that the order 1 expansion of  $\lambda$  actually vanishes. By contrast, recall that the corresponding term did not vanish in the derivation of the Itô–vector projection which lead us to give an alternative derivation using the weak Itô–Taylor expansion.

**COROLLARY 4.8.** *Let  $K$  be a compact set in  $\mathbb{R}^r$  with  $M \cup K$  closed and containing a neighbourhood of  $X_0$ . Let  $d^{g^E, K}$  be the semimetric defined by (3.1) where we take  $g^E$  to be the Euclidean metric. Let  $d^{g, K}$  be the semimetric defined by taking  $g$  to be the Riemannian metric induced on  $M$ .*

*Let  $Y$  is a process on  $M$  given locally by (4.2), and  $X$  is the process given by (4.1). If we assume both these equations have smooth coefficients then*

$$E(d^K(X_t, \phi(Y_t))^2) = Ct^2 + O(t^{\frac{5}{2}})$$

*for some constant  $C$  if and only if equation (4.10) holds. Moreover the value of  $C$  is then minimized by taking  $A$  as defined in (4.11). Note that this choice is independent of  $K$ .*

*The same holds for  $d^{g, K}$  provided the SDEs are assumed to be autonomous (or if one is willing to assume the generalization of Proposition 3.2).*

*Proof.* For  $d^K$ , this is an application of Proposition 3.1 together with the argument of Corollary 4.3. For  $d^{g, K}$ , this is an application of Proposition 3.2.  $\square$

We finish by proving a simply differential geometry Lemma that was used in the proof of Theorem 4.7.

**LEMMA 4.9.** *Let  $U$  be a neighbourhood of the origin in  $\mathbb{R}^n$  and let  $\phi : U \rightarrow \mathbb{R}^r$  be normal coordinates for the Riemannian manifold  $\phi(U)$  centred at the origin, then*

$$|\phi(x) - \phi(y)|_r^2 = |x - y|_n^2 + O((|x|_n + |y|_n)^4).$$

*Here  $|\cdot|_n$  is the norm on  $\mathbb{R}^n$ .*

*Proof.* Without loss of generality we may assume that the origin is mapped to the origin and the coordinate axes in  $\mathbb{R}^n$  are mapped to the corresponding axes in  $\mathbb{R}^r$ . Given a point  $y \in U$ , we can write the Taylor expansion for the component  $\phi(y)^a$  in the following form:

$$\phi(y)^a = (\delta_n^r)_i^a y^i + A_{jk}^a y^j y^k + O(|y|_n^3). \quad (4.15)$$

Here  $(\delta_n^r)$  is the tensor representing the projection of  $\mathbb{R}^r$  onto  $\mathbb{R}^n$ . The upper indices of  $(\delta_n^r)$  range from 1 to  $r$  and the lower from 1 to  $n$ .  $(\delta_n^r)_j^i$  is equal to 1 if  $i = j$  and 0 otherwise.  $A_{jk}^a$  is a tensor with upper index  $a$  ranging from 1 to  $r$  and lower indices  $j$  and  $k$  ranging from 1 to  $n$  and which satisfies  $A_{jk}^a = A_{kj}^a$ .

The components of the metric tensor on  $U$  can now be computed as follows:

$$\begin{aligned} g_{ij} &= \left\langle \frac{\partial \phi}{\partial y^i}, \frac{\partial \phi}{\partial y^j} \right\rangle_r \\ &= ((\delta_n^r)_i^a + A_{ik}^a y^k) ((\delta_n^r)_j^b + A_{jl}^b y^l) (g^r)_{ab} + O(|y|_n^2) \end{aligned}$$

Here  $g^r$  is the metric tensor of  $\mathbb{R}^r$ . Our expression for  $g_{ij}$  simplifies to give:

$$g_{ij} = (g^n)_{ij} + A^j_{ik}y^k + A^i_{jk}y^k + O(|y|_n^2).$$

It is well known that in Riemannian normal coordinates the partial derivatives of the metric tensor vanish at the origin. We compute that

$$\partial_k g_{ij}|_0 = A^j_{ik} + A^i_{jk}.$$

So we have  $A^j_{ik} = -A^i_{jk}$ . However, recall that  $A^i_{jk}$  is symmetric in the indices  $j$  and  $k$ . We see that:

$$A^i_{jk} = -A^j_{ik} = -A^j_{ki} = A^k_{ji} = A^k_{ij} = -A^i_{kj} = -A^i_{jk}.$$

So all the components of  $A$  vanish.

We can now use (4.15) to compute:

$$\begin{aligned} |\phi(x) - \phi(y)|_r^2 &= |(\delta_n^r)_i^a x^i - (\delta_n^r)_i^a y^i|_r^2 + O((|x|_n + |y|_n)^4) \\ &= |x - y|_n^2 + O((|x|_n + |y|_n)^4). \end{aligned}$$

□

#### 4.4. Time-symmetric optimality of the Stratonovich projection

We now consider the question of whether it is possible to derive the Stratonovich projection via an optimality argument. We show this is possible, but our construction uses time-symmetric optimality criteria. This would be an unnatural choice for most applications.

The optimality criteria we have used earlier for the Itô projections minimize the mean square of the Taylor expansion of the difference between the solution of two SDEs at a positive time  $t$  given that they have the same initial state 0. These criteria are asymmetric under time reflections, as positive times have a special status. This explains why the Itô projections that we have derived are asymmetrical in time. By contrast, Stratonovich calculus is symmetric under reflections in time, so it is perhaps not so surprising that these optimality criteria do not yield the Stratonovich projection. We therefore ask what a time-symmetric optimality criterion should look like.

Consider the  $x$  SDE (4.5), in Stratonovich form:

$$dx_t = \bar{a}(x, t)dt + b_\alpha(x, t) \circ dW_t^\alpha, \quad X_0., \quad (4.16)$$

where  $\bar{a}$  is given by the Itô Stratonovich transformation

$$\bar{a}_i = a_i - \frac{1}{2} \sum_{j=1}^{d_x} \sum_{\alpha} b_\alpha^j \frac{\partial b_\alpha^i}{\partial x_j}.$$

More generally, by a bar over the drift of an Itô SDE we will mean the drift of the equivalent Stratonovich SDE.

We extend the SDE to negative time as follows. Define

$$d\xi_t = -\bar{a}(\xi, t)dt - b_\alpha(\xi, t) \circ d\hat{W}_t^\alpha, \quad \xi_0 = X_0 \quad (4.17)$$

where  $\hat{W}$  is a second standard Brownian motion, independent of  $W$ . Given the symmetric nature of the Stratonovich integral underlying the above SDE and given that formally the chain rule holds, it makes sense to define  $x$  for  $t < 0$  by setting

$$x_{-t} := \xi_t.$$

We now wonder whether the Stratonovich projection could be indeed optimal at time 0 for this SDE extended to negative time at time 0. Suppose that we wish to find the SDE on  $M$

$$dy = \bar{A} dt + B_\alpha \circ dW_t^\alpha, \quad y_0 = x_0, \quad (4.18)$$

extended similarly to negative time (giving  $y_{-t}$ ), that minimizes the mean square of the truncated Taylor expansion of the vector  $(f(x_t) - F(y_t)) \oplus (f(x_{-t}) - F(y_{-t}))$ . Here  $f$  and  $F$  are functions as defined in Proposition 4.1.

We show in [2] that the Stratonovich projection arises from this optimality criterion.

Nevertheless, in typical applications of SDEs such as stochastic filtering, there is a clear time-asymmetry modelled by a filtration of the probability space. For such applications, the second time-reversed Brownian motion would be irrelevant to the problem, and hence a time-symmetric optimality criterion would be inappropriate.

### 5. A low dimensional example: cross diffusion on a unit circle

We now look at a concrete example which shows the difference between the Itô-vector and Itô-jet projections. Consider the SDE in  $\mathbb{R}^2$  given by

$$\begin{aligned} dX_t &= \sigma Y_t dW_t, \\ dY_t &= \sigma X_t dW_t, \end{aligned} \tag{5.1}$$

with deterministic initial condition  $(X_0, Y_0)$ . We call this a cross diffusion, since each state crosses over as diffusion coefficient of the other state and the paths tend to lie on a St Andrew cross, see the appendix in [2] for more details on this process. We wish to project this process equation onto the unit circle given by  $X^2 + Y^2 = 1$ . It is easy to check using Itô's Lemma that if we write  $(X_t, Y_t)$  in polar coordinates as  $(r_t \cos(\theta_t), r_t \sin(\theta_t))$  then  $\theta_t = \arctan(Y_t/X_t)$  satisfies the following exact angular position process equation:

$$d\theta_t = -\frac{1}{2}\sigma^2 \sin(4\theta_t)dt + \sigma \cos(2\theta_t)dW_t, \quad \text{or} \quad d\theta_t = \sigma \cos(2\theta_t) \circ dW_t. \tag{5.2}$$

Thanks to the special structure of the cross-diffusion, the equation above is already a closed SDE for  $\theta$  without needing to apply any of our projection methods. In this sense we already have the exact angular position SDE and we do not need to project the original  $\mathbb{R}^2$  SDE on the circle  $M$  to approximate the exact angular position with a SDE on the circle. However, we might want to check whether one of our projection methods is consistent with the exact angular position SDE. Let us check how the different projections behave. If we use the same polar coordinate  $\theta$  for the unit circle, we find that the Stratonovich projection and the Itô-jet projection for the  $(X, Y)$  SDE are also given by (5.2), and are thus consistent with the exact  $\theta$ . However the Itô-vector projection is different and results in:

$$d\theta_t = \sigma \cos(2\theta_t)dW_t.$$

For this example at least, the Itô-jet projection and the Stratonovich projections track the angular position of  $(X_t, Y_t)$  perfectly. Intuitively one might therefore feel that the Stratonovich and Itô-jet projections are “better” approximations to the SDE despite the short time optimality arguments given earlier. It turns out this is a special case of a more general situation, summarized in the following

**DEFINITION 9** SDE that fibers over a map between manifolds. Let  $f : M \rightarrow N$  be a smooth map between two manifolds. Let  $S$  be an SDE on  $M$  determined by the 2-jets  $\gamma_x : \mathbb{R}^m \rightarrow M$  given at each point  $x \in M$ . We say that  $S$  *fibres over*  $f$  if  $j_2(f \circ \gamma_{x_1}) = j_2(f \circ \gamma_{x_2})$  whenever  $f(x_1) = f(x_2)$ . This implies that we can define an SDE on the image of  $f$  using the 2-jets  $j_2(f \circ \gamma_x)$  at  $f(x)$ . We call this the *SDE induced by*  $f$ .

Returning to projection, we see that we have the following

**THEOREM 5.1** (If SDE fibres over  $\Pi^s$  then Stratonovich = Itô-jet proj.). *If an SDE fibres over the smooth projection map  $\Pi^s$  then the Stratonovich and Itô-jet projection will both be equal to the SDE induced by  $\Pi^s$ .*

*Proof.* This is an immediate consequence of the Stratonovich chain rule in the first case. It is a trivial consequence of the definition of the Itô-jet projection in the second case.  $\square$

Our two-dimensional example of the cross-diffusion on the circle is simply a special case of this more general phenomenon.

It is interesting to note that one can draw a diagram to show the Itô-jet projection. In [3] it is discussed how the jet formulation of SDEs makes it possible to draw pictures of SDEs that transform according to Itô's lemma. For processes driven by one dimensional Brownian motion, one simply finds functions  $\gamma_x$  whose 2-jet represents the SDE and then draws the image of an interval  $[-\epsilon, \epsilon]$  under the map  $\gamma_x$  at each point  $x$ . A picture of this type is shown in Figure 4. It shows how the 2-jets determining the SDE (5.1) can be projected onto the unit circle simply by composition with  $\Pi^s$ .

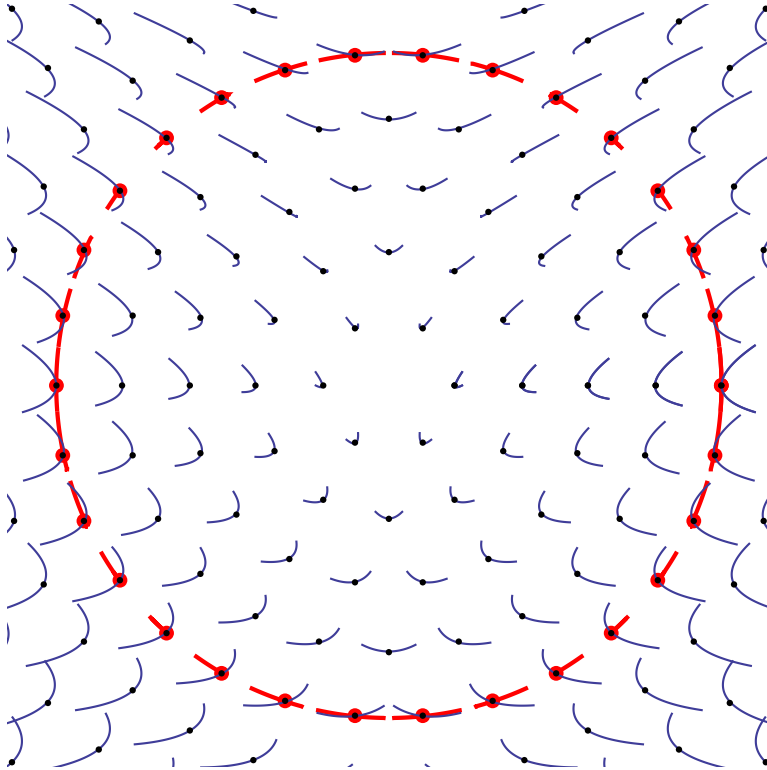


FIGURE 4. An SDE in  $\mathbb{R}^2$  and its Itô-jet projection onto the unit circle

It seems paradoxical that we derived the Itô-vector projection using optimality arguments that seem to be less ad hoc than for the Stratonovich projection, and yet, for this example, the Itô-vector projection appears manifestly suboptimal.

One possible resolution to this paradox is to say that our notions of tracking  $X_t$  optimally are flawed. Theorem 4.2 has the weakness that we attempt to minimize a term of order 1 when our approximation is not accurate at order  $\frac{1}{2}$ . Indeed, looking at equation (4.6) we see that

when we try to minimize the relevant expectation we minimize a combination of terms of order  $t$  and  $t^2$  for the square. Moreover, Theorem 4.4 has the weakness that we are using the error in the mean to measure the accuracy of our solution. By contrast, the Itô-jet projection has a fully convincing derivation as the optimal approximation of  $\Pi^s(X_t)$  up to order 1.

We will see numerical evidence later that suggests that the Itô-jet projection performs better in the long term than the Itô-vector projection which lends some support to the idea that the Itô-jet projection is the “right” choice.

We summarize the different projections and the optimality criteria used to determine their drifts in Table 1. The diffusion coefficient is identical for all three projections.

Projection	Properties of drift term
Itô-vector	Minimizes absolute value of order 1 Itô–Taylor expansion for $E(X - \phi(Y))$ .
Itô-jet	Minimizes order 2 Itô–Taylor expansion for expected $\mathbb{R}^r$ or $M$ distance between $\Pi^s(X)$ & $\phi(Y)$ .
Stratonovich	Similar to Itô-jet above but for the Taylor series of the differences vector $[X_t - \phi(Y_t), X_{-t} - \phi(Y_{-t})]$ at positive and negative time, where negative time processes are defined ad hoc by propagating a second input Brownian motion backward in time.

TABLE 1. *Projections and the associated optimality criteria*

### 6. The Itô-jet projection in local coordinates

Our definition of the Itô-jet projection is coordinate free and simple. However, to calculate it in practice we will need an explicit coordinate representation.

We therefore wish to calculate the metric projection map  $\tilde{\Pi}^s = \psi \circ \Pi^s$  up to second order. Then using Itô’s formula for 2-jets we will be able to calculate the Itô-jet projection associated to  $\tilde{\Pi}^s$ .

Most of our calculation involves the deterministic map  $\tilde{\Pi}^s$ . Thus in this section we will drop the convention of using Greek indices exclusively for components of the Brownian motion. In this section we will also use Greek indices to highlight indices over which we are summing. This makes the formulae a little easier to read.

We define the metric tensor on  $U$  by:

$$h_{ab} = \frac{\partial \phi^\alpha}{\partial x^a} \frac{\partial \phi^\alpha}{\partial x^b} \quad (6.1)$$

The differential  $\tilde{\Pi}_*^s$  of  $\tilde{\Pi}^s$  is well known to be given by the linear projection onto  $\text{Im } \phi_*$  composed with the map  $\phi_*^{-1}$ . Hence  $\tilde{\Pi}_*^s$  is the unique linear map with  $\tilde{\Pi}_*^s \circ \phi_*$  equal to the identity and with kernel equal to the orthogonal complement of  $\text{Im } \phi_*$ . We deduce that  $\tilde{\Pi}_*^s$  has the following components:

$$\Pi_b^a := (\tilde{\Pi}_*^s)_b^a = \frac{\partial \phi^b}{\partial x^a} h^{a\alpha}, \quad a \leq n, \alpha \leq n, b \leq r. \quad (6.2)$$

We note that the differential or tangent map  $\tilde{\Pi}_*^s$  is the best linear approximation of the metric projection  $\tilde{\Pi}^s$  around the relevant point  $x = \phi(y) \in M$ , and it coincides with the classic linear projection  $\Pi_{\phi(y)}$  on the tangent space of  $M$ . Indeed, equation (6.2) shows the classic components of the projection on the tangent space of an  $n$ -dimensional manifold  $M$  embedded in  $\mathbb{R}^r$  and realized as  $\phi$ -image of a subset of  $\mathbb{R}^n$ .



LEMMA 6.1. *Suppose for simplicity that  $\phi(0) = 0$  and*

$$(\phi_*)_b^a := \frac{\partial \phi^a}{\partial x^b} = D_b^a := \begin{cases} 1 & a = b \text{ and } a \leq n \\ 0 & \text{otherwise} \end{cases}$$

then  $\Pi^s$  is given up to second order by

$$\tilde{\Pi}^s(y)^a = y^a - \frac{1}{2} \frac{\partial^2 \phi^a}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta + \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} (\perp)_{\gamma\alpha} y^\alpha y^\beta + O(|y|^3)$$

where we define

$$(\perp)_{ab} = \begin{cases} 1 & a = b \text{ and } a > n \\ 0 & \text{otherwise.} \end{cases}$$

Note that we are using an extension of the Einstein summation convention to cover tensors where some indices range from 1 to  $n$  and some from 1 to  $r$ . Where an index appears twice, we sum over the smaller range. Note also that we are working in a restricted set of coordinate systems, so it no longer holds that all summed pairs of indices will consist of an upper and a lower index.

*Proof.* By our simplifying assumption we may write:

$$\tilde{\Pi}^s(y)^a = y^a + A_{\alpha\beta}^a y^\alpha y^\beta + B_{\alpha\beta\gamma}^a y^\alpha y^\beta y^\gamma + O(|y|^4) \quad (6.3)$$

where  $A_{\alpha\beta}^a$  is symmetric in  $\alpha$  and  $\beta$  and  $B$  is symmetric in  $\alpha, \beta$  and  $\gamma$ . The Taylor series expansion for  $\phi$  now allows us to compute the components of  $(y - \phi(\tilde{\Pi}^s(y)))$ .

$$\begin{aligned} (y - \phi(\tilde{\Pi}^s(y)))^a &= y^a - D_\alpha^a (y^\alpha + A_{\beta\gamma}^\alpha y^\beta y^\gamma + B_{\beta\gamma\delta}^\alpha y^\beta y^\gamma y^\delta) \\ &\quad - \frac{1}{2} \frac{\partial^2 \phi^a}{\partial x^\alpha \partial x^\beta} (y^\alpha + A_{\delta\epsilon}^\alpha y^\delta y^\epsilon) (y^\beta + A_{\zeta\eta}^\beta y^\zeta y^\eta) + O(|y|^4) \\ &= y^a - D_\alpha^a y^\alpha \\ &\quad - D_\alpha^a A_{\beta\gamma}^\alpha y^\beta y^\gamma - \frac{1}{2} \frac{\partial^2 \phi^a}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta \\ &\quad - \frac{\partial^2 \phi^a}{\partial x^\alpha \partial x^\beta} A_{\delta\epsilon}^\alpha y^\delta y^\epsilon y^\beta - D_\alpha^a B_{\beta\gamma\delta}^\alpha y^\beta y^\gamma y^\delta + O(|y|^4) \end{aligned}$$

We take the partial derivative of this with respect to  $A_{qr}^p$  to get:

$$\frac{\partial}{\partial A_{qr}^p} (y - \phi(\tilde{\Pi}^s(y)))^a = -D_p^a y^q y^r - \frac{\partial^2 \phi^a}{\partial x^p \partial x^\beta} y^q y^r y^\beta + O(|y|^4).$$

Because of the distance minimizing property of  $\Pi^s$  we know that for all  $p, q, r$  and sufficiently small  $y$  we have:

$$\frac{\partial}{\partial A_{qr}^p} |(y - \phi(\tilde{\Pi}^s(y)))|^2 = 0$$

The left hand side of this expression is equal to:

$$2 \left( \frac{\partial}{\partial A_{qr}^p} (y - \phi(\tilde{\Pi}^s(y)))^a \right) (y - \phi(\tilde{\Pi}^s(y)))^a.$$

We have written down explicit expressions for each term in this product. This enables us to write down the fourth order terms of  $\frac{\partial}{\partial A_{qr}^p} |(y - \phi(\tilde{\Pi}^s(y)))|^2$ . They are given by:

$$\begin{aligned} & 2D_p^a A_{\alpha\beta}^a y^\alpha y^\beta y^q y^r + D_p^a \frac{\partial^2 \phi^a}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta y^q y^r - 2 \frac{\partial^2 \phi^a}{\partial x^p \partial x^\beta} (y^a - D_\gamma^a y^\gamma) y^\beta y^q y^r \\ &= \left( 2A_{\alpha\beta}^p y^\alpha y^\beta + \frac{\partial^2 \phi^p}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta - 2 \frac{\partial^2 \phi^a}{\partial x^p \partial x^\beta} (y^a - D_\delta^a y^\delta) y^\beta \right) y^q y^r \\ &= \left( 2A_{\alpha\beta}^p y^\alpha y^\beta + \frac{\partial^2 \phi^p}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta - 2 \frac{\partial^2 \phi^a}{\partial x^p \partial x^\beta} (\perp)_{\alpha a} y^\alpha y^\beta \right) y^q y^r \end{aligned}$$

We know that this must vanish for all sufficiently small  $y$ . We deduce that

$$2A_{\alpha\beta}^p y^\alpha y^\beta + \frac{\partial^2 \phi^p}{\partial x^\alpha \partial x^\beta} y^\alpha y^\beta - 2 \frac{\partial^2 \phi^a}{\partial x^p \partial x^\beta} (\perp)_{\alpha a} y^\alpha y^\beta = 0.$$

for all sufficiently small  $y$ . This gives us an expression for  $A_{\alpha\beta}^p y^\alpha y^\beta$  which combines with equation (6.3) to prove the result.  $\square$

We now use the lemma coupled with some coordinate transformations to compute a second order expression for the metric projection in the general case.

**PROPOSITION 6.2.** *Let  $g_\perp^\phi$  be the symmetric two form on  $\mathbb{R}^r$  defined by:*

$$g_\perp^\phi(X + X^\perp, Y + Y^\perp) = g(X^\perp, Y^\perp) \quad X, Y \in \text{Im } \phi \text{ and } X^\perp, Y^\perp \in (\text{Im } \phi)^\perp$$

where  $g$  is the Euclidean metric on  $\mathbb{R}^r$ . Define coordinates  $\tilde{y}$  centered on  $0 \in R^r$  by  $\tilde{y}^a = y^a - y_0^a$ . Then to second order the metric projection is given by

$$\begin{aligned} \tilde{\Pi}^s(y)^a &= x_0 + \Pi_\alpha^a \tilde{y}^\alpha - \frac{1}{2} \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\gamma^\alpha \Pi_\delta^\alpha \Pi_\epsilon^\beta \tilde{y}^\delta \tilde{y}^\epsilon \\ &\quad + \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\delta^\beta h^{a\alpha} \tilde{y}^\gamma \tilde{y}^\delta - \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\epsilon^\beta \Pi_\gamma^\eta \Pi_\delta^\zeta h_{\eta\zeta} h^{a\alpha} \tilde{y}^\delta \tilde{y}^\epsilon \\ &\quad + O(|y|^3) \end{aligned}$$

and where  $\Pi$  is given by equation (6.2).

*Proof.* We assume without loss of generality that  $x_0 = 0$  and  $y_0 = \phi(x_0) = 0$ .

We can find a coordinate transformation  $J$  of  $\mathbb{R}^n$  which maps an orthonormal basis of  $\mathbb{R}^n$  to the standard basis vectors. We take  $x$  to be our original coordinates and  $X$  to be the coordinates obtained by applying  $J^{-1}$ . So we have:

$$x^a = J_b^a X^b$$

To satisfy our requirements  $J$  must satisfy:

$$h_{\alpha\beta} J_a^\alpha J_b^\beta = \delta_{ab}$$

Equivalently:

$$h_{ab} = (J^{-1})_a^\alpha (J^{-1})_b^\alpha.$$

So any pseudo square root of  $h_{ab}$  will give an appropriate choice for  $J^{-1}$ . Taking the matrix inverse of the above expression we have:

$$h^{ab} = J_a^a J_a^b \quad (6.4)$$

We can now find an orthogonal transformation  $T$  of  $\mathbb{R}^r$  mapping  $\text{Im } \phi^*$  to  $\mathbb{R}^n \subseteq \mathbb{R}^r$ . Hence  $\Phi = T \circ \phi \circ J$  satisfies  $(\Phi^*)_b^a = D_b^a$ . We will write  $x$  for the original coordinates on  $\mathbb{R}^r$  and

define transformed coordinates  $X$  by:

$$X^a = T_\alpha^a x^\alpha.$$

Let us write  $\Pi'$  for the metric projection associated with the map  $\Phi$ . The various maps we have just defined are summarized in the commutative diagram below:

$$\begin{array}{ccc} \mathbb{R}^n & \begin{array}{c} \xrightarrow{\Phi} \\ \xleftarrow{\Pi'} \end{array} & \mathbb{R}^r \\ J \downarrow & & \uparrow T \\ \mathbb{R}^n & \begin{array}{c} \xrightarrow{\phi} \\ \xleftarrow{\tilde{\Pi}^s} \end{array} & \mathbb{R}^r \end{array}$$

From Lemma 6.1 we have:

$$\begin{aligned} \Pi'(Y)^a &= \frac{\partial \Phi^a}{\partial X^\alpha} Y^\alpha - \frac{1}{2} \frac{\partial^2 \Phi^a}{\partial X^\alpha \partial X^\beta} Y^\alpha Y^\beta \\ &\quad + \frac{\partial^2 \Phi^\gamma}{\partial X^a \partial X^\beta} (\perp)_{\gamma\alpha} Y^\alpha Y^\beta + O(|Y|^3) \\ &= \frac{\partial(T^{-1} \circ \Phi)^\beta}{\partial X^\alpha} T_\beta^a Y^\alpha - \frac{1}{2} \frac{\partial^2(T^{-1} \circ \Phi)^\gamma}{\partial X^\alpha \partial X^\beta} T_\gamma^a Y^\alpha Y^\beta \\ &\quad + \frac{\partial^2(T^{-1} \circ \Phi)^\delta}{\partial X^a \partial X^\beta} T_\delta^\gamma (\perp)_{\gamma\alpha} Y^\alpha Y^\beta + O(|Y|^3). \end{aligned}$$

Hence

$$\begin{aligned} \Pi'(Ty)^a &= \frac{\partial(T^{-1} \circ \Phi)^\beta}{\partial X^\alpha} T_\beta^a T_\gamma^\alpha y^\gamma - \frac{1}{2} \frac{\partial^2(T^{-1} \circ \Phi)^\gamma}{\partial X^\alpha \partial X^\beta} T_\gamma^a T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta \\ &\quad + \frac{\partial^2(T^{-1} \circ \Phi)^\delta}{\partial X^a \partial X^\beta} T_\delta^\gamma (\perp)_{\gamma\alpha} T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta + O(|y|^3) \\ &= \frac{\partial(T^{-1} \circ \Phi)^\beta}{\partial x^\gamma} J_\alpha^\gamma T_\beta^a T_\gamma^\alpha y^\gamma - \frac{1}{2} \frac{\partial^2(T^{-1} \circ \Phi)^\gamma}{\partial x^\iota \partial x^\kappa} J_\alpha^\iota J_\beta^\kappa T_\gamma^a T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta \\ &\quad + \frac{\partial^2(T^{-1} \circ \Phi)^\delta}{\partial x^\iota \partial x^\kappa} J_\alpha^\iota J_\beta^\kappa T_\delta^\gamma (\perp)_{\gamma\alpha} T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta + O(|y|^3). \end{aligned}$$

We deduce that:

$$\begin{aligned} \tilde{\Pi}^s(y)^a &= ((J \circ \pi' \circ T)(y))^a \\ &= \frac{\partial \phi^\beta}{\partial x^\delta} J_\lambda^a J_\alpha^\delta T_\beta^\lambda T_\gamma^\alpha y^\gamma - \frac{1}{2} \frac{\partial^2 \phi^\gamma}{\partial x^\iota \partial x^\kappa} J_\lambda^\iota J_\alpha^\kappa J_\beta^\lambda T_\gamma^a T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta \\ &\quad + \frac{\partial^2 \phi^\delta}{\partial x^\iota \partial x^\kappa} J_\lambda^\iota J_\alpha^\kappa J_\beta^\lambda T_\delta^\gamma (\perp)_{\gamma\alpha} T_\zeta^\alpha T_\eta^\beta y^\zeta y^\eta + O(|y|^3). \end{aligned}$$

We now note that:

$$\Pi_b^a = J_\alpha^a D_\beta^\alpha T_b^\beta = J_\alpha^a T_b^\alpha.$$

This allows us to simplify our expression for  $\Pi^s(y)$  to:

$$\begin{aligned}\tilde{\Pi}^s(y)^a &= \frac{\partial\phi^\beta}{\partial x^\delta} \Pi_\beta^a \Pi_\gamma^\delta y^\gamma - \frac{1}{2} \frac{\partial^2\phi^\gamma}{\partial x^\iota \partial x^\kappa} \Pi_\gamma^a \Pi_\zeta^\iota \Pi_\eta^\kappa y^\zeta y^\eta \\ &\quad + \frac{\partial^2\phi^\delta}{\partial x^\iota \partial x^\kappa} \Pi_\eta^\kappa J_\lambda^\iota J_\lambda^\kappa T_\delta^\gamma(\perp)_{\gamma\alpha} T_\zeta^\alpha y^\zeta y^\eta + O(|y|^3).\end{aligned}$$

The tensor  $\perp_{ab}$  is equal to  $(g_\perp^\Phi)_{ab}$ . So since  $T$  is an isometry, we may write

$$(\perp)_{\alpha\beta} T_a^\alpha T_b^\beta = (g_\perp^\Phi)_{ab}.$$

Using this together with equation (6.4) we may write:

$$\begin{aligned}\tilde{\Pi}^s(y)^a &= \frac{\partial\phi^\beta}{\partial x^\delta} \Pi_\beta^a \Pi_\gamma^\delta y^\gamma - \frac{1}{2} \frac{\partial^2\phi^\gamma}{\partial x^\iota \partial x^\kappa} \Pi_\gamma^a \Pi_\zeta^\iota \Pi_\eta^\kappa y^\zeta y^\eta \\ &\quad + \frac{\partial^2\phi^\delta}{\partial x^\iota \partial x^\kappa} \Pi_\eta^\kappa h^{a\iota} (g_\perp^\Phi)_{\delta\zeta} y^\zeta y^\eta + O(|y|^3)\end{aligned}$$

The first term can be simplified by repeated applications of equations (6.1) and (6.2):

$$\begin{aligned}\frac{\partial\phi^\beta}{\partial x^\delta} \Pi_\beta^a \Pi_\gamma^\delta y^\gamma &= \frac{\partial\phi^\beta}{\partial x^\delta} \frac{\partial\phi^\beta}{\partial x^\gamma} h^{a\gamma} \frac{\partial\phi^\gamma}{\partial x^\epsilon} h^{\epsilon\delta} y^\gamma \\ &= h_{\delta\gamma} h^{a\gamma} \frac{\partial\phi^\gamma}{\partial x^\epsilon} h^{\epsilon\delta} y^\gamma = \frac{\partial\phi^\gamma}{\partial x^\epsilon} h^{\epsilon a} y^\gamma = \Pi_\beta^a y^\beta.\end{aligned}$$

It is a tautology that the first order term is given by  $\Pi$ , nevertheless this calculation is a reassuring check on our working. Renaming the dummy variables we now have that:

$$\begin{aligned}\tilde{\Pi}^s(y)^a &= x_0 + \Pi_\alpha^a y^\alpha - \frac{1}{2} \frac{\partial^2\phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\gamma^a \Pi_\delta^\alpha \Pi_\epsilon^\beta y^\delta y^\epsilon \\ &\quad + \frac{\partial^2\phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\epsilon^\beta h^{a\alpha} (g_\perp^\Phi)_{\gamma\delta} y^\delta y^\epsilon + O(|y|^3)\end{aligned}$$

We would like a formula that can be computed efficiently when  $n \ll r$ , so we wish to eliminate the term  $g_\perp^\Phi$ . By splitting vectors  $V$  and  $W$  in  $\mathbb{R}^r$  into components in  $\text{Im } \phi_*$  and its orthogonal complement, we see that the Euclidean metric on  $\mathbb{R}^r$  satisfies the decomposition:

$$g_{ab} V^a W^b = (g_\perp^\Phi)_{ab} V^a W^b + h_{ab} \Pi_\alpha^a V^\alpha \Pi_\alpha^b W^\beta.$$

Using this formula we obtain:

$$\begin{aligned}\tilde{\Pi}^s(y)^a &= x_0 + \Pi_\alpha^a y^\alpha - \frac{1}{2} \frac{\partial^2\phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\gamma^a \Pi_\delta^\alpha \Pi_\epsilon^\beta y^\delta y^\epsilon \\ &\quad + \frac{\partial^2\phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\delta^\beta h^{a\alpha} y^\gamma y^\delta - \frac{\partial^2\phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\epsilon^\beta \Pi_\gamma^\eta \Pi_\delta^\zeta h_{\eta\zeta} h^{a\alpha} y^\delta y^\epsilon \\ &\quad + O(|y|^3)\end{aligned}$$

□

We can immediately conclude:

**THEOREM 6.3** (Itô-jet projection in coordinates). *Let  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^r$  be an embedding with  $\phi(x_0) = y_0$  then the Itô-jet projection of the SDE:*

$$dy = a dt + b_\alpha dW_t^\alpha, \quad y_0$$

is

$$dx = A dt + B_\alpha dW_t^\alpha, \quad x_0$$

where:

$$B_\alpha^i = \Pi_\beta^i b_\alpha^\beta$$

and:

$$\begin{aligned} A^i &= \Pi_\alpha^i a^\alpha + \\ &\quad \left( -\frac{1}{2} \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\gamma^i \Pi_\delta^\alpha \Pi_\epsilon^\beta \right. \\ &\quad \left. + \frac{\partial^2 \phi^\epsilon}{\partial x^\alpha \partial x^\beta} \Pi_\delta^\beta h^{i\alpha} - \frac{\partial^2 \phi^\gamma}{\partial x^\alpha \partial x^\beta} \Pi_\epsilon^\beta \Pi_\gamma^\eta \Pi_\delta^\zeta h_{\eta\zeta} h^{i\alpha} \right) \\ &\quad \times b_\kappa^\delta b_\iota^\epsilon [W^\kappa, W^\iota]_t. \end{aligned}$$

$\Pi$  is given by (6.2).  $h_{ab}$  is given by (6.1).  $h^{ab}$  is the inverse of  $h_{ab}$ .

It is reassuring to check that this formula gives the same result as we found in Section 5 for projection of a particular SDE onto a circle where the projection map was known exactly. In fact, we can find an explicit expression for the Itô-jet projection of any bivariate SDE driven by a single Brownian motion on the plane on the unit circle.

EXAMPLE 1 Itô-jet projection of a bivariate SDE on the unit circle. Suppose that our diffusion process in  $\mathbb{R}^r = \mathbb{R}^2$ , driven by a one-dimensional Brownian motion  $W = W^1$ , is

$$dX = a_1(X, Y)dt + b_1^1(X, Y)dW^1, \quad X_0$$

$$dY = a_2(X, Y)dt + b_1^2(X, Y)dW^1, \quad Y_0$$

and suppose we wish to approximate this process in the unit circle. If we define  $\theta = \arctan(Y_t/X_t)$ , and compute  $d\theta_t$  via Itô's formula, this won't be in general a closed SDE for  $\theta$ , contrary to the special example of the cross diffusion above. To obtain a closed SDE in  $\theta$  we have to project. One can check that for the one-dimensional manifold given by the unit circle, expressed as

$$M = \{(\cos(\theta), \sin(\theta)), \theta \in [0, 2\pi)\}$$

with coordinates  $Y = \theta$  in  $\mathbb{R}^n = \mathbb{R}^1$ , one has

$$h = 1, \quad h^{-1} = 1, \quad \Pi_1 = -\sin(\theta), \quad \Pi_2 = \cos(\theta), \quad \partial_\theta^2 \phi^1 = -\cos(\theta), \quad \partial_\theta^2 \phi^2 = -\sin(\theta),$$

which allows us to apply Theorem 6.3 to this system. We obtain (coefficients  $a$  and  $b$  are computed in  $X = \cos(\theta), Y = \sin(\theta)$ )

$$A(\theta) = -a_1 \sin(\theta) + a_2 \cos(\theta) + \frac{1}{2} \sin(2\theta)((b_1^1)^2 - (b_1^2)^2) - \cos(2\theta)b_1^1 b_1^2,$$

$$B(\theta) = -\sin(\theta)b_1^1 + \cos(\theta)b_1^2.$$

In the special case of  $a_1 = a_2 = 0$  and  $b_1^1 = \sigma \sin(\theta), b_1^2 = \sigma \cos(\theta)$  this confirms our previous calculations for the cross-diffusion example.

## 7. Application of the Projection to Non-linear Filtering

As a fundamental application of our new projection methods we consider an area from signal processing, stochastic filtering. This extends our previous work in [7]. For geometric methods in stochastic filtering in general we refer to [16].

In stochastic filtering one has a signal  $X$  that evolves according to a SDE, and observes a process  $Y$  which is a function of this signal plus noise. This is standard notation, but these  $X$  and  $Y$  are not to be confused with the processes we used earlier in the paper, in that they are not the  $\mathbb{R}^r$  process to be approximated and its  $\mathbb{R}^n$  approximation.

The filtering problem consists in estimating the signal  $X$  given the present and past observations  $Y$ . If  $t$  is the current time, the solution of the filtering problem is the probability density of the state  $X_t$  conditional on the observations from time 0 to time  $t$ , call it  $p_t$ . The density  $p_t$  follows the Kushner-Stratonovich (or Zakai) stochastic partial differential equation (SPDE) that, under some technical assumptions, can be seen as a stochastic differential equation in the infinite dimensional  $L^2$  space of square roots of densities (Hellinger metric) or of densities themselves (direct  $L^2$  metric).

The process we wish to approximate on a low dimensional manifold is  $p_t$ , which represents the  $X_t$  of our earlier sections. The  $\mathbb{R}^r$  space of our earlier sections is the  $L^2$  infinite dimensional space, while the submanifold  $M$  is a finite dimensional family of probability densities parametrized by  $\theta$ , acting as coordinates:  $\{p(\cdot, \theta), \theta \in \Theta \subset \mathbb{R}^n\}$ .  $\theta_t$  plays the role of what we were calling  $Y_t$  earlier in the paper. We aim at finding a SDE for  $\theta$  such that  $p(\cdot, \theta_t)$  approximates  $p_t(\cdot)$  in an optimal way. Note that in the previous part of the paper we had a dimensionality reduction from  $r$  to  $n$ , whereas now we go from infinite dimensional  $p_t$  to  $n$ -dimensional  $\theta_t$ .

One may be concerned about taking our finite dimensional results and applying them in an infinite dimensional setting. However, we have stated our results in terms of approximating one Ito–Taylor series of a given order with another Ito–Taylor series. This allows us to avoid the analytical issues that might conceivably arise in considering the convergence of these series. Therefore our results generalize straightforwardly to the Hilbert space setting. As an example, the minimization argument used to prove Theorem 4.2 relies only on properties of the linear projection operator that remain true in a Hilbert space setting.

In addition the explicit calculation of Section 6 can be generalized unproblematically to the case of a finite dimensional manifold embedded in a Hilbert space. To see this simply note that the vector space spanned by the first two derivatives of the map  $\phi$  at  $p$  gives a finite dimensional space  $V$  and so one can simply apply the result for embedding into the space  $V$ .

The point where complexities might conceivably arise in the infinite dimensional setting is in the generalizations of Proposition 2.2 and Proposition 2.3. Folk wisdom suggests that such results can be generalized to Hilbert spaces without difficulty, so we will not attempt to prove that here.

### 7.1. The Kushner Stratonovich equation

We suppose that the state  $X_t \in \mathbb{R}^n$  of a system evolves according to the equation:

$$dX_t = f(X_t, t) dt + \sigma(X_t, t) dW_t$$

where  $f$  and  $\sigma$  are smooth  $\mathbb{R}^n$  valued functions and  $W_t$  is a Brownian motion. One typically adds growth conditions to ensure a global existence and uniqueness result for the signal equation, see for example [7] and references therein for the details.

We suppose that an associated process, the observation process,  $Y_t \in \mathbb{R}^d$  evolves according to the equation:

$$dY_t = b(X_t, t) dt + dV_t$$

where  $b$  is a smooth  $\mathbb{R}^d$  valued function and  $V_t$  is a Brownian motion independent of  $W_t$ . Note that the filtering problem is often formulated with an additional constant in terms of the observation noise. For simplicity we have assumed that the system is scaled so that this can be omitted.

The filtering problem is to compute the conditional distribution of  $X_t$  given a prior distribution for  $X_0$  and the values of  $Y$  for all times up to and including  $t$ .

Subject to various bounds on the growth of the coefficients of this equation, the assumption that the distribution has a density  $p_t$  and suitable bounds on the growth of  $p_t$  one can show

that  $p_t$  satisfies the Kushner–Stratonovich SPDE:

$$dp = \mathcal{L}^* p \, dt + p[b - E_p(b)]^T [dY - E_p(b)dt] \quad (7.1)$$

where  $E_p$  denotes the expectation with respect to the density  $p$ ,

$E_p[f] = \int f(x)p(x)dx$ , and the forward diffusion operator  $\mathcal{L}_t^*$  is defined by:

$$\mathcal{L}_t^* \phi = -\frac{\partial}{\partial x^i} [f_i(x, t)\phi] + \frac{1}{2} \frac{\partial^2}{\partial x^i \partial x^j} [a_{ij}(x, t)\phi] \quad (7.2)$$

where  $a = \sigma\sigma^T$ . Note that we are using the Einstein summation convention in this expression.

In the event that the coefficient functions  $f$  and  $b$  are all linear and  $\sigma$  is a deterministic function of time one can show that so long as the prior distribution for  $X$  is Gaussian, or deterministic, the density  $p$  will be Gaussian at all subsequent times. This allows one to reduce the infinite dimensional equation (7.1) to a finite dimensional stochastic differential equation for the mean and covariance matrix of this normal distribution. This finite dimensional problem solution is known as the Kalman filter.

For more general coefficient functions, however, equation (7.1) cannot be reduced to a finite dimensional problem [20]. Instead one might seek approximate solutions of (7.1) that belong to some given statistical family of densities. This is a very general setup and includes, for example, approximating the density using piecewise linear functions to derive a finite difference approximation or approximating the density with Hermite polynomials to derive a spectral method. Other examples include exponential families (considered in [11, 10]) and mixture families (considered in [5, 7]).

Our projection theory tells us how one can find good approximations on a given statistical family with respect to a given metric on the space of distributions. We illustrate this by writing down the Itô-vector and Itô-jet projection of (7.1) for the  $L^2$  and Hellinger metrics on a general manifold<sup>†</sup>.

We will then examine some numerical results regarding the very specific case of seeking approximate solutions using Gaussian distributions. The idea of approximating the solution to the filtering problem using a Gaussian distribution has been considered by numerous authors who have derived variously, the extended Kalman filter [28], assumed density filters [25] and Stratonovich projection filters [10]. Some of these are related, for example the assumed density filters and Stratonovich projection filters in Hellinger metrics for Gaussian (and more generally exponential) families coincide [11]. Using our new projection methods, we will be able to derive projection filters which outperform all these other filters (assuming performance is measured over small time intervals using the appropriate Hilbert space metric).

We note that (7.1) is an infinite dimensional SDE driven by a continuous semi-martingale. The definitions and results given in Section 2 were only stated in the finite dimensional case for SDEs driven by Brownian motion. The definition of Itô–Taylor series can be generalized straightforwardly to this situation and hence the definition of the Itô projections can be applied in this context also.

More generally, for the the geometry of infinite dimensional filtering problems based on  $L^2$  or Orlicz charts and for the related differential geometric approach to statistics with recent advances we refer for example to [29, 26, 27, 18, 11, 7, 12, 13].

## 7.2. Itô-vector projections

**7.2.1. The Itô-vector projection filter in the  $L^2$  direct metric** Let us suppose that the density  $p$  lies in  $L^2$  and so we can use the  $L^2$  norm to measure the accuracy of an approximate

---

<sup>†</sup>Note that it is also possible to consider projecting the Zakai equation. However, as explained in [7], one expects that projecting the Kushner–Stratonovich will lead to smaller error terms.

solution to equation (7.1). For a discussion on conditions under which a unnormalized version of  $p$  is in  $L^2$  (Zakai Equation) see for example [1].

We wish to consider an  $m$ -dimensional family of distributions  $p$  parameterized by  $m$  real valued parameters  $\theta^1, \theta^2, \dots, \theta^m$ . For example we will consider the 2 dimensional Gaussian family:

$$p(x) = \frac{1}{(\theta^2)\sqrt{2\pi}} \exp\left(-\frac{(x - (\theta^1))^2}{2(\theta^2)^2}\right). \quad (7.3)$$

Note that we have chosen to follow differential geometry convention and use upper indices for the coordinate functions  $\theta^i$  so we have been careful to distinguish powers from indices using brackets.

More formally, an  $m$ -dimensional family is given by a smooth embedding  $\phi : \mathbb{R}^m \rightarrow L^2(\mathbb{R}^n)$ . The tangent vectors  $\phi_* \frac{\partial}{\partial \theta^i} \in L^2(\mathbb{R}^n)$  are simply the partial derivatives

$$\frac{\partial p}{\partial \theta^i}.$$

Let us write:

$$g_{ij} = \int_{\mathbb{R}} \frac{\partial p}{\partial \theta^i} \frac{\partial p}{\partial \theta^j} dx.$$

This defines the induced metric tensor on the manifold  $\phi(\mathbb{R}^m)$ . We will write  $g^{ij}$  for the inverse of the matrix  $g_{ij}$ . The projection operator  $\Pi_{\phi(\theta)}$  is then given by

$$\begin{aligned} \Pi_{\phi(\theta)}(v) &= \sum_{i,j=1}^m g^{ij} \left\langle v, \phi_* \frac{\partial}{\partial \theta^i} \right\rangle_{L^2} \phi_* \frac{\partial}{\partial \theta^j} \\ &= \sum_{i,j=1}^m g^{ij} \left( \int_{\mathbb{R}^n} v(x) \frac{\partial p}{\partial \theta^i} dx \right) \phi_* \frac{\partial}{\partial \theta^j}. \end{aligned}$$

Thus

$$\phi_*^{-1} \Pi_{\phi(\theta)}(v) = \sum_{i,j=1}^m g^{ij} \left( \int_{\mathbb{R}^n} v(x) \frac{\partial p}{\partial \theta^i} dx \right) \frac{\partial}{\partial \theta^j}.$$

We can now write down the Itô-vector projection of (7.1) with respect to the  $L^2$  metric. It is:

$$d\theta^i = A^i dt + B^i dY_t$$

where:

$$B^i = \sum_{j=1}^m g^{ij} \left( \int_{\mathbb{R}} (p(b - E_{p(\theta)}(b)))^T \frac{\partial p}{\partial \theta^j} dx \right)$$

and

$$A^i = \sum_{j=1}^m g^{ij} \left( \int_{\mathbb{R}^n} \left( \mathcal{L}^* p - p(b - E_{p(\theta)}(b))^T E_{p(\theta)}(b) - \frac{1}{2} \sum_{k=1}^m \frac{\partial^2 p}{\partial \theta^j \partial \theta^k} B^k \right) \frac{\partial p}{\partial \theta^j} dx \right).$$

**EXAMPLE 2** Itô-vector projection filter for cubic sensor in direct metric. Consider as a test case the 1-dimensional problem with  $f(x, t) = 0$ ,  $\sigma(x, t) = 1$  and  $b(x, t) = x + \epsilon x^3$  for some small constant  $\epsilon$ . This problem is a perturbation of a linear filter so one might expect that a Gaussian approximation will perform reasonably well at least for small times. Thus we will use the 2 dimensional manifold of Gaussian distributions given in equation (7.3).

We first calculate the metric tensor  $g_{ij}$  which is diagonal in this case:

$$g_{ij} = \frac{1}{4\sqrt{\pi}(\theta^2)^3} \begin{pmatrix} 1 & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$



This is easily inverted to compute  $g^{ij}$ . We compute the expectation  $E_p(b)$ :

$$E_p(b) = \frac{\epsilon (\sqrt{2\pi}(\theta^1)^3(\theta^2) + 3\sqrt{2\pi}(\theta^1)(\theta^2)^3)}{\sqrt{2\pi}(\theta^2)} + (\theta^1).$$

One can now see that computing the projection equation will simply involve integrating a number of terms of the form a polynomial in  $x$  times a Gaussian. The end result is:

$$\begin{aligned} d\theta^1 &= \left( -\frac{1}{4}\theta^1(\theta^2)^2 \left( 3\epsilon^2 \left( 4(\theta^1)^4 - 4(\theta^2)^2(\theta^1)^2 - 3(\theta^2)^4 \right) + 16\epsilon(\theta^1)^2 + 4 \right) \right) dt \\ &\quad + \left( \frac{1}{2}(\theta^2)^2 \left( 3\epsilon \left( 2(\theta^1)^2 + (\theta^2)^2 \right) + 2 \right) \right) dY_t \\ d\theta^2 &= \left( -\frac{9\epsilon^2(\theta^2)^8 + (\theta^2)^4 \left( 60\epsilon^2(\theta^1)^4 + 48\epsilon(\theta^1)^2 + 4 \right) + 6\epsilon(\theta^2)^6 \left( 9\epsilon(\theta^1)^2 + 2 \right) - 4}{8\theta^2} \right) dt \\ &\quad + \left( 3\epsilon\theta^1(\theta^2)^3 \right) dY_t \end{aligned}$$

**7.2.2. The Itô-vector projection filter in the Hellinger metric** The Hellinger metric is a metric on probability measures. In the case of two probability density functions  $p(x)$  and  $q(x)$  on  $\mathbb{R}^n$ , that now need only be in  $L^1$ , the Hellinger distance is given by the square root of:

$$\frac{1}{2} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx.$$

In other words, up to the constant factor of  $\frac{1}{2}$  the Hellinger metric corresponds to the  $L^2$  norm on the square root of the density function rather than on the density itself (as in the previous subsection). The Hellinger metric has the important advantage of making the metric independent of the particular background density that is used to express measures as densities. The  $L^2$  direct distance introduced earlier does not satisfy this background independence.

Now, to compute the Itô-vector projection with respect to the Hellinger metric we first want to write down an Itô equation for the evolution on  $\sqrt{p}$ .

Applying Itô's lemma to equation (7.1) we formally obtain:

$$\begin{aligned} d\sqrt{p} &= \left( \frac{\mathcal{L}^*p - p(b - E_p(b))^T E_p(b)}{2\sqrt{p}} - \frac{p^2(b - E_p(b))^T(b - E_p(b))}{8p\sqrt{p}} \right) dt \\ &\quad + \left( \frac{p(b - E_p(b))^T}{2\sqrt{p}} \right) dY_t. \\ &= \left( \frac{\mathcal{L}^*p}{2\sqrt{p}} - \frac{1}{8}\sqrt{p}(b - E_p(b))^T(b + 3E_p(b)) \right) dt \\ &\quad + \left( \frac{1}{2}\sqrt{p}(b - E_p(b))^T \right) dY_t. \end{aligned}$$

A family of distributions now corresponds to an embedding  $\phi$  from  $\mathbb{R}^m$  to  $L^2(\mathbb{R}^n)$  but now  $p = \phi(\theta)^2$ . The tangent space is spanned by the vectors:

$$\phi_* \frac{\partial}{\partial \theta^i} = \frac{\partial \sqrt{p}}{\partial \theta^i}.$$

We define a metric on the tangent space by:

$$h_{ij} = \int_{\mathbb{R}^n} \frac{\partial \sqrt{p}}{\partial \theta^i} \frac{\partial \sqrt{p}}{\partial \theta^j} dx.$$

We write  $h^{ij}$  for the inverse matrix of  $h_{ij}$ . The projection operator with respect to the Hellinger metric is:

$$\Pi_{\phi(\theta)}(v) = \sum_{i,j=1}^m h^{ij} \left( \int_{\mathbb{R}^n} v(x) \frac{\partial \sqrt{p}}{\partial \theta^i} dx \right) \phi_* \frac{\partial}{\partial \theta^j}.$$

We can now write down the Itô-vector projection of (7.1) with respect to the Hellinger metric. It is:

$$d\theta^i = A^i dt + B^i dY_t$$

where:

$$B^i = \sum_{j=1}^m h^{ij} \left( \int_{\mathbb{R}} \frac{1}{2} \sqrt{p}(b - E_{p(\theta)}(b))^T \frac{\partial \sqrt{p}}{\partial \theta^j} dx \right)$$

and

$$A^i = \sum_{j=1}^m h^{ij} \left( \int_{\mathbb{R}^n} \left( \frac{\mathcal{L}^* p}{2\sqrt{p}} - \frac{1}{8} \sqrt{p}(b - E_{p(\theta)}(b))^T (b + 3E_{p(\theta)}(b)) \right. \right. \\ \left. \left. - \frac{1}{2} \sum_{k=1}^m \frac{\partial^2 \sqrt{p}}{\partial \theta^j \partial \theta^k} B^k \right) \frac{\partial \sqrt{p}}{\partial \theta^j} dx \right).$$

EXAMPLE 3 Itô-vector projection filter for cubic sensor: Hellinger metric. We may repeat Example 2 but projecting using the Hellinger metric. We first calculate the metric tensor  $h_{ij}$  which is diagonal also in this case:

$$h_{ij} = \frac{1}{4\theta_2^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

This is easily inverted to compute  $h^{ij}$ . We obtain the following SDEs:

$$d\theta^1 = \left( -\theta^1 (\theta^2)^2 \left( 3\epsilon^2 \left( (\theta^1)^4 + 4(\theta^2)^2 (\theta^1)^2 + 6(\theta^2)^4 \right) + \epsilon \left( 4(\theta^1)^2 + 6(\theta^2)^2 \right) + 1 \right) \right. \\ \left. + \left( (\theta^2)^2 \left( 3\epsilon \left( (\theta^1)^2 + (\theta^2)^2 \right) + 1 \right) \right) dY_t \right) dt \\ d\theta^2 = \left( -\frac{27\epsilon^2 (\theta^2)^8 + (\theta^2)^4 (15\epsilon^2 (\theta^1)^4 + 12\epsilon (\theta^1)^2 + 1) + 9\epsilon (\theta^2)^6 (6\epsilon (\theta^1)^2 + 1) - 1}{2\theta^2} \right) dt \\ \left. + \left( 3\epsilon \theta^1 (\theta^2)^3 \right) dY_t \right)$$

### 7.3. Itô-jet projections

Using the formulae from Theorem 6.3 together with the formulae and techniques of Section 7.2 we can explicitly calculate the Itô-vector projections of the filtering equation in both the  $L^2$  and Hellinger metrics.

To minimize notation, let us concentrate on the 1-dimensional state space filtering problem and project using the  $L^2$  metric.

We can formally write the filtering equation in the form:

$$dp_t = \mu(p_t)dt + \Sigma(p_t)dW_t \quad (7.4)$$

where  $p_t$  is an  $L^2$  function and

$$\mu(p)(x) := \frac{1}{2} \frac{d^2(\sigma(x)^2 p(x))}{dx^2} - \frac{d(f(x)p(x))}{dx} \\ - p(x) \left( b(x) - \int_{\mathbb{R}} p(t)b(t)dt \right) \int_{\mathbb{R}} p(t)b(t)dt, \quad (7.5) \\ \Sigma(p)(x) := p(x) \left( b(x) - \int_{\mathbb{R}} p(t)b(t)dt \right).$$

We now suppose that  $p_t$  is parameterized as  $p_t(x) = \phi(\theta)(x)$  as in Section 7.2. Using Theorem 6.3 we can write down the Itô-jet projection which is an SDE for the components of  $\theta$ .

To write down the result it will be useful to define functions  $\pi^i(\theta)$  by:

$$\pi^i(\theta) = h^{ij} \frac{\partial \phi}{\partial \theta^j}(\theta).$$

We will also use angle brackets to denote the  $L^2$  inner product. With this understood, the Itô-jet projection of the filtering equations in the  $L^2$  metric is given by:

$$d\theta_t^i = A^i(\theta)dt + B^i(\theta)dW_t$$

where we have in turn

$$B^i(\theta) = \langle \pi^i(\theta), \Sigma(\phi(\theta)) \rangle dt$$

and

$$\begin{aligned} A^i(\theta) = & \langle \pi^i(\theta), \mu(\phi(\theta)) \rangle \\ & - \frac{1}{2} \left\langle \frac{\partial^2 \phi}{\partial \theta^\alpha \partial \theta^\beta}(\theta), \pi^i(\theta) \right\rangle \langle \Sigma(\theta), \pi^\alpha(\theta) \rangle \langle \Sigma(\theta), \pi^\beta(\theta) \rangle \\ & + \left\langle \frac{\partial^2 \phi}{\partial \theta^\alpha \partial \theta^\beta}(\theta), \Sigma(\phi(\theta)) \right\rangle \langle \pi^\beta(\theta), \Sigma(\phi(\theta)) \rangle h^{i\alpha}(\theta) \\ & - \left\langle \frac{\partial^2 \phi}{\partial \theta^\alpha \partial \theta^\beta}, \pi^\eta(\theta) \right\rangle \langle \pi^\beta(\theta), \Sigma(\phi(\theta)) \rangle \langle \pi^\xi(\theta), \Sigma(\phi(\theta)) \rangle h_{\eta\xi}(\theta) h^{i\alpha}(\theta) \end{aligned}$$

EXAMPLE 4 Itô-jet projection filter for cubic sensor in direct metric. For the filtering problem of Example 2 the Itô-jet projection in the  $L^2$  metric is

$$\begin{aligned} d\theta^1 = & \left( -\frac{1}{4}\theta^1(\theta^2)^2 \left( 3\epsilon^2 \left( 4(\theta^1)^4 - 4(\theta^2)^2(\theta^1)^2 - 9(\theta^2)^4 \right) + 16\epsilon(\theta^1)^2 + 4 \right) \right) dt \\ & + \left( \frac{1}{2}(\theta^2)^2 \left( 3\epsilon \left( 2(\theta^1)^2 + (\theta^2)^2 \right) + 2 \right) \right) dY_t \\ d\theta^2 = & \left( \frac{3\epsilon^2(\theta^2)^8 - 4(\theta^2)^4(15\epsilon^2(\theta^1)^4 + 12\epsilon(\theta^1)^2 + 1) - 2\epsilon(\theta^2)^6(15\epsilon(\theta^1)^2 + 2) + 4}{8\theta^2} \right) dt \\ & + \left( 3\epsilon\theta^1(\theta^2)^3 \right) dY_t \end{aligned}$$

The Itô-jet projection of the filtering equation in the Hellinger metric can be computed in the same way. Indeed we can formally write the filtering equation in the form:

$$dq_t = \mu(q_t)dt + \Sigma(q_t)dW_t \tag{7.6}$$

where  $q_t$  is the square root of the density and the coefficients now satisfy

$$\begin{aligned} \mu(q)(x) := & \frac{1}{2q(x)} \left( \frac{1}{2} \frac{d^2(\sigma(x)^2 q(x)^2)}{dx^2} - \frac{d(f(x)q(x)^2)}{dx} \right) \\ & - \frac{1}{8} q(x) \left( b(x) - \int_{\mathbb{R}} q(t)^2 b(t) dt \right) \left( b(x) + 3 \int_{\mathbb{R}} q(t)^2 b(t) dt \right), \\ \Sigma(q)(x) := & \frac{1}{2} q(x) \left( b(x) - \int_{\mathbb{R}} q(t)^2 b(t) dt \right). \end{aligned} \tag{7.7}$$

Thus we can use the same formulae as above to compute the Hellinger projection except we must use the coefficients from (7.7) rather than those from (7.5).

EXAMPLE 5 Itô-jet projection filter for cubic sensor: Hellinger metric. For the filtering problem of Example 2, the Itô-jet projection in the Hellinger metric is

$$\begin{aligned} d\theta^1 &= \left( -\theta^1 (\theta^2)^2 \left( 3\epsilon^2 \left( (\theta^1)^4 + 4(\theta^2)^2 (\theta^1)^2 + 3(\theta^2)^4 \right) + \epsilon \left( 4(\theta^1)^2 + 6(\theta^2)^2 \right) + 1 \right) \right) dt \\ &\quad + \left( (\theta^2)^2 \left( 3\epsilon \left( (\theta^1)^2 + (\theta^2)^2 \right) + 1 \right) \right) dY_t \\ d\theta^2 &= \left( -\frac{18\epsilon^2 (\theta^2)^8 + (\theta^2)^4 (15\epsilon^2 (\theta^1)^4 + 12\epsilon (\theta^1)^2 + 1) + 3\epsilon (\theta^2)^6 (15\epsilon (\theta^1)^2 + 2) - 1}{2\theta^2} \right) dt \\ &\quad + \left( 3\epsilon\theta^1 (\theta^2)^3 \right) dY_t \end{aligned}$$

In [2] and [6] we detail the different classic Gaussian approximate filters we will use to test our newly derived filters. These include the Stratonovich projection filter, the assumed density filters and the extended Kalman filter.

#### 7.4. Results

Our explicit calculations show that the two Itô projections give rise to new, distinct, Gaussian approximations.

All our calculations of the resulting filters for the cubic sensor  $b(x, t) = x + \epsilon x^3$  are equal when  $\epsilon = 0$ . This provides a basic sanity check that our formulae correspond to the Kalman filter in the case of a linear sensor. In general, if we know that the solution lies in a particular manifold and we project onto that manifold, the three projection methods will all be exact.

We simulated the example problem  $b(x) = x + \epsilon x^3$  for all of the above approximate filters with  $\epsilon = 0.05$ . We also computed an “exact” solution using a finite difference method on a grid of 1000 intervals spaced evenly from  $-10.0$  to  $10.0$  and a time step of  $0.0002$ . We define the  $L^2$  residual to be the  $L^2$  distance between the approximate solution and the “exact” solution. We define the Hellinger residual similarly, as the  $L^2$  distance between the square roots of the solution densities.

In Figure 5 we see the  $L^2$  residuals for the various methods. All the projection methods shown are taken using the  $L^2$  metric in this case. The Itô-vector projection in the  $L^2$  metric results in the lowest residuals over short time horizons. The Stratonovich projection comes a close second. Over medium term time horizons, the Itô-jet projection out performs the Itô-vector projection. We have not shown longer term behaviour because over long time horizons, all the methods become inaccurate and any comparison becomes meaningless. The projection methods out-performed all other methods. Although our plot shows only a single run, it is reasonably representative of the typical behaviour.

In Figure 6 we have plotted the ratio of the Hellinger residual for each method to the residual of the Itô-jet projection w.r.t. the Hellinger metric. This is because the residuals themselves are too difficult to distinguish visually. Thus values exceeding 1 show a larger error than the Itô-jet projection and values less than one show a lower error. All the projection methods shown in this plot are taken w.r.t. the Hellinger metric.

This plot indicates that the Itô ADF and the Itô-jet projection are almost indistinguishable in their performance. A look at the explicit formulae reveals that the difference between these two equations is of order  $\epsilon^2$  whereas the difference between the other equations is of order only  $\epsilon$ . Over the short term, the Itô-vector projection gives the best results. Over the medium term, the Itô-jet projection and the Itô ADF give the best results. Again, over the longer term all the filters become highly inaccurate.

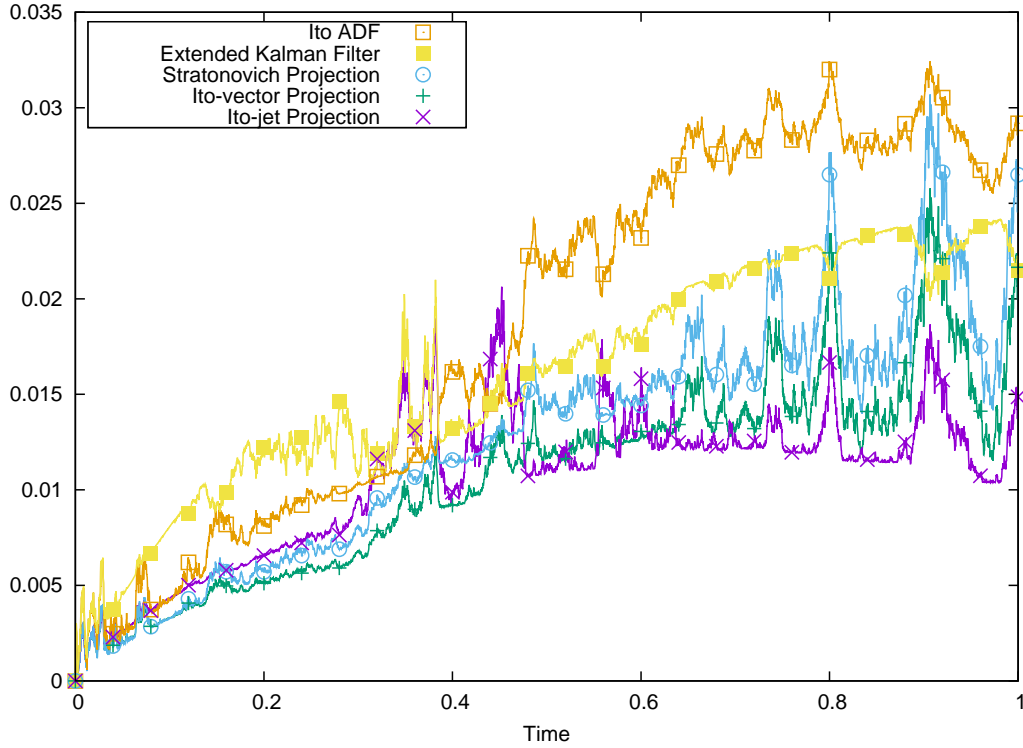


FIGURE 5.  $L^2$  residuals for each approximation method. All projections are taken relative to the  $L^2$  metric.

## 8. Conclusions

The notion of projecting a vector field onto a manifold is unambiguous. By contrast, there are multiple distinct generalizations of this notion to SDEs, as summarized in Table 1.

The two Itô projections we introduced in this work can both be derived from minimization arguments. However, the Itô-jet projection has some clear advantages.

- The Itô-jet projection is the best approximation to the metric projection of the true solution and has an error of  $O(t)$ . By contrast the Itô-vector projection only tracks the true solution an accuracy of  $O(t^{\frac{1}{2}})$ .
- The Itô-jet projection gives a more intuitive answer than the Itô-vector projection for the low dimensional example considered in Section 5.
- The Itô-jet projection gives better numerical results in the medium term than the Itô-vector projection in our application to filtering.
- The Itô-jet projection has an elegant definition when written in terms of 2-jets.
- The Itô-jet projection has a pictorial interpretation, shown in Figure 4.

We have also seen that the Stratonovich projection satisfies an ad hoc minimization that is less appealing than the ones of the Itô projections, since it requires a deterministic anchor point. The Itô-jet and Itô-vector projection arguments allow one to derive new Gaussian approximations to non-linear filters. Unlike previous Gaussian approximations to non-linear filters, these approximations are derived by minimization arguments rather than heuristic arguments. Thus the notion of projecting an SDE onto a manifold is able to give new results even for this well-worn topic.

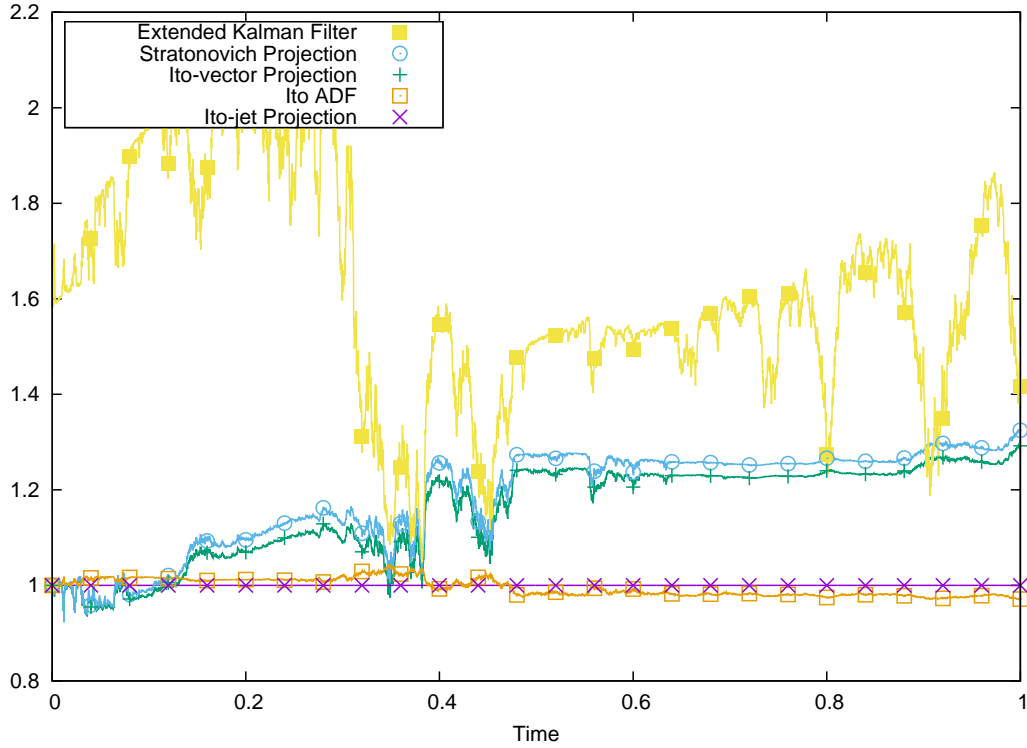


FIGURE 6. Hellinger residuals for various approximation method divided by residual for the Itô-jet projection. All projections are taken relative to the Hellinger metric.

### References

1. N. U. Ahmed. *Linear and Nonlinear Filtering for Scientists and Engineers*. World Scientific, Singapore, 1998.
2. J. Armstrong and D. Brigo. Optimal approximation of SDEs on submanifolds: the Ito vector and Ito jet projections. <https://arxiv.org/abs/1610.03887>, 2016.
3. J. Armstrong and D. Brigo. Intrinsic stochastic differential equations as jets. *Proc. R. Soc. A*, 474(2210), 2018.
4. J. Armstrong, D. Brigo, and E. Rossi Ferrucci. Projections of SDEs onto Submanifolds. *ArXiv preprint arXiv:1810.03923*, 2018.
5. John Armstrong and Damiano Brigo. Stochastic filtering via L2 projection on mixture manifolds with computer algorithms and numerical examples. *arXiv preprint arXiv:1303.6236*, 2013.
6. John Armstrong and Damiano Brigo. Extrinsic projection of Itô SDEs on submanifolds with applications to non-linear filtering. In: Nielsen, F., Critchley, F., & Dodson, K. (Eds), *Computational Information Geometry for Image and Signal Processing*, Springer Verlag, 2016.
7. John Armstrong and Damiano Brigo. Nonlinear filtering via stochastic PDE projection on mixture manifolds in  $L^2$  direct metric. *Mathematics of Control, Signals and Systems*, 28(1):1–33, 2016.
8. Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*, volume 3. Springer, 2009.
9. L. C. Brewin. Riemann normal coordinates. *Preprint Department of Mathematics, Monash University, Clayton, Victoria*, 3168, 1997.
10. Damiano Brigo, Bernard Hanzon, and François LeGland. A differential geometric approach to nonlinear filtering: the projection filter. *Automatic Control, IEEE Transactions on*, 43(2):247–252, 1998.
11. Damiano Brigo, Bernard Hanzon, and François LeGland. Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli*, 5(3):495–534, 1999.
12. Damiano Brigo and Giovanni Pistone. Dimensionality reduction for measure-valued evolution equations in statistical manifolds. In *Proceedings of the conference on Computational Information Geometry for Image and Signal Processing*. Springer, 2016.
13. Damiano Brigo and Giovanni Pistone. Optimal approximations of the Fokker-Planck-Kolmogorov equation: projection, maximum likelihood eigenfunctions and Galerkin methods. *arXiv preprint arXiv:1603.04348*, 2016.

14. Z. Brzeźniak and K. D. Elworthy. Stochastic differential equations on Banach manifolds. *Methods Funct. Anal. Topology*, 6(1):43–84, 2000.
15. David Elworthy. Geometric aspects of diffusions on manifolds. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 277–425. Springer, 1988.
16. David Elworthy, Yves Le Jan, and Xue-Mei Li. *The Geometry of Filtering*. Frontiers in Mathematics. Birkhauser, 2010.
17. M. Emery. *Stochastic calculus in manifolds*. Springer-Verlag, Heidelberg, 1989.
18. Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.
19. Yuri E. Gliklikh. *Global and Stochastic Analysis with Applications to Mathematical Physics*. Theoretical and Mathematical Physics. Springer, London, 2011.
20. Michiel Hazewinkel, SI Marcus, and HJ Sussmann. Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem. *Systems & control letters*, 3(6):331–340, 1983.
21. Elton P. Hsu. *Stochastic Analysis on Manifolds*. Contemporary Mathematics. American Mathematical Society, 2002.
22. K. Itô. Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.*, (1):35–47, 1950.
23. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
24. Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*. Applications and Mathematics. Springer, Berlin, New York, Third printing, 1999.
25. Harold J Kushner. Approximations to optimal nonlinear filters. *Automatic Control, IEEE Transactions on*, 12(5):546–556, 1967.
26. Nigel J. Newton. An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.*, 263(6):1661–1681, 2012.
27. Nigel J. Newton. Information geometric nonlinear filtering. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 2(18):1550014, 24, 2015.
28. Étienne Pardoux. Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *Ecole d'Été de Probabilités de Saint-Flour XIX 1989*, pages 68–163. Springer, 1991.
29. Giovanni Pistone and Carlo Sempì. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*, 23(5):1543–1561, 10 1995.
30. L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

John Armstrong  
 Dept. of Mathematics  
 King's College London  
 WC2R 2LS London, UK  
 john.1.armstrong@kcl.ac.uk

Damiano Brigo, Emilio Rossi Ferrucci  
 Dept. of Mathematics  
 Imperial College London  
 SW7 2AZ London, UK  
 damiano.brigo@imperial.ac.uk,  
 emilio.rossi-ferrucci16@imperial.ac.uk